

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 8, 2003, 09:36:57 ; Search time 43 Seconds
(without alignments)
6072.583 Million cell updates/sec

Title: US-09-556-246-1

Perfect score: 7526

Sequence: 1 MAMKTLPIYLLLLSVFVIQ.....ARAITRSGQTLSKWNVNC 1404

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 684280 seqs, 185983659 residues

Total number of hits satisfying chosen parameters: 684280

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications AA.*
1: /cgn2_6/ptodata/2/pubppa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubppa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubppa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubppa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubppa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubppa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubppa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubppa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubppa/US09_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubppa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubppa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubppa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubppa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubppa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubppa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubppa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubppa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubppa/US60_PUBCOMB.pep.*

Prod. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7526	100.0	1404	US-10-124-557-2	Sequence 2, Appl1
2	7526	100.0	1404	US-10-124-557-62	Sequence 62, Appl1
3	7291.5	96.9	1361	US-10-124-557-40	Sequence 40, Appl1
4	7245.5	96.3	1363	US-10-124-557-52	Sequence 52, Appl1
5	7232	96.1	1354	US-10-124-557-48	Sequence 48, Appl1
6	7011	93.2	1320	US-10-124-557-46	Sequence 46, Appl1
7	7011	93.2	1320	US-10-124-557-60	Sequence 60, Appl1
8	7007.5	93.1	1311	US-10-124-557-42	Sequence 42, Appl1
9	6971	92.6	1314	US-10-124-557-50	Sequence 50, Appl1
10	6951.5	92.4	1313	US-10-124-557-142	Sequence 142, Appl1
11	6737	89.4	1270	US-10-124-557-44	Sequence 44, Appl1
12	6085	80.9	1140	US-10-124-557-104	Sequence 104, Appl1
13	5510.5	73.2	1049	US-10-124-557-58	Sequence 58, Appl1
14	5425.5	72.1	1022	US-10-124-557-84	Sequence 84, Appl1
15	5335	70.9	1038	US-10-124-557-74	Sequence 74, Appl1

16	5011	66.6	941	US-10-124-557-14	Sequence 14, Appl1
17	2850.5	37.9	538	US-10-038-694-3	Sequence 3, Appl1
18	2034.5	27.0	463	US-10-124-557-64	Sequence 54, Appl1
19	1763.5	23.4	423	US-10-124-557-66	Sequence 66, Appl1
20	1754	23.3	422	US-10-124-557-68	Sequence 64, Appl1
21	1460	19.4	372	US-10-124-557-64	Sequence 96, Appl1
22	1188	15.8	220	US-10-124-557-96	Sequence 106, Appl1
23	1188	15.8	5179	US-09-922-217-1068	Sequence 1068, Appl1
24	1188	15.8	5179	US-09-833-263-1068	Sequence 1068, Appl1
25	1188	15.8	5179	US-10-023-380-1068	Sequence 94, Appl1
26	1126	15.0	209	US-10-124-557-94	Sequence 116, Appl1
27	1121	14.9	208	US-10-124-557-132	Sequence 92, Appl1
28	1111.5	14.8	207	US-10-124-557-136	Sequence 70, Appl1
29	1111.5	14.8	207	US-10-124-557-136	Sequence 30, Appl1
30	1100	14.6	204	US-10-124-557-92	Sequence 90, Appl1
31	1077	14.3	296	US-10-124-557-70	Sequence 88, Appl1
32	1074	14.3	231	US-10-124-557-30	Sequence 32, Appl1
33	1041	13.8	192	US-10-124-557-10	Sequence 106, Appl1
34	968.5	12.9	188	US-10-038-694-2	Sequence 118, Appl1
35	945	12.6	172	US-10-124-557-88	Sequence 102, Appl1
36	881	11.7	188	US-10-124-557-32	Sequence 114, Appl1
37	871	11.6	156	US-10-124-557-106	Sequence 108, Appl1
38	871	11.6	156	US-10-124-557-118	Sequence 76, Appl1
39	827.5	11.0	157	US-10-124-557-118	Sequence 80, Appl1
40	827.5	11.0	157	US-10-124-557-118	Sequence 82, Appl1
41	809	10.7	1367	US-09-801-368-108	Sequence 14, Appl1
42	806	10.7	217	US-10-124-557-76	Sequence 76, Appl1
43	805	10.7	1325	US-09-864-761-35612	Sequence 35612, A
44	800	10.6	141	US-10-124-557-80	Sequence 80, Appl1
45	750	10.0	131	US-10-124-557-82	Sequence 82, Appl1

ALIGNMENTS

RESULT 1
US-10-124-557-2
Sequence 2, Application US/10124557
Publication No. US20020137894A1
GENERAL INFORMATION:
APPLICANT: Turner, Katherine
Clark, Stephen C.
Jacob, Kenneth
Hewick, Rodney M.
Geener, Thomas G.
NUMBER OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/124,557
FILING DATE: 16-Apr-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:

NAME: Caeir, Luann
 REGISTRATION NUMBER: 31,822
 REFERENCE/DOCKET NUMBER: GI 5190
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (617) 876-1170
 TELEFAX: (617) 876-5851
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1404 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 SEQUENCE DESCRIPTION: SEQ ID NO: 2:
 US-10-124-557-2

Query Match 100.0%; Score 7526; DB 14; Length 1404;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1404; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 MAMKTLPIYLILLISVFIYQVSSQDLSGACGCEGVSROATTCNCDYNCQHYMECCPDF 60
DB 1 MAMKTLPIYLILLISVFIYQVSSQDLSGACGCEGVSROATTCNCDYNCQHYMECCPDF 60
QY 61 KRVTAEISCKRCESFEGRGCCDAQCKKYDKCCPDYESFCAEVHNPTSPSSKKAP 120
DB 61 KRVTAEISCKRCESFEGRGCCDAQCKKYDKCCPDYESFCAEVHNPTSPSSKKAP 120
QY 61 KRVTAEISCKRCESFEGRGCCDAQCKKYDKCCPDYESFCAEVHNPTSPSSKKAP 120
DB 61 KRVTAEISCKRCESFEGRGCCDAQCKKYDKCCPDYESFCAEVHNPTSPSSKKAP 120
QY 121 PPSGASQITKSTTKSPKPPNKKTKYVIESEITEEHSVSNQSSSSSSSSSSSTTW 180
DB 121 PPSGASQITKSTTKSPKPPNKKTKYVIESEITEEHSVSNQSSSSSSSSSSSTTW 180
QY 121 PPSGASQITKSTTKSPKPPNKKTKYVIESEITEEHSVSNQSSSSSSSSSSSTTW 180
DB 121 PPSGASQITKSTTKSPKPPNKKTKYVIESEITEEHSVSNQSSSSSSSSSSSTTW 180
QY 181 KIKSSKSNANRELOKLVKQDKNKKRKKKTPKPVVDEAGSLDNDGDFVTTTDPST 240
DB 181 KIKSSKSNANRELOKLVKQDKNKKRKKKTPKPVVDEAGSLDNDGDFVTTTDPST 240
QY 181 KIKSSKSNANRELOKLVKQDKNKKRKKKTPKPVVDEAGSLDNDGDFVTTTDPST 240
DB 181 KIKSSKSNANRELOKLVKQDKNKKRKKKTPKPVVDEAGSLDNDGDFVTTTDPST 240
QY 241 TQHNKVSSTPKITTAAPINRPSLPNSDTSKETSITVANKETVTEKETTNNKQSTDG 300
DB 241 TQHNKVSSTPKITTAAPINRPSLPNSDTSKETSITVANKETVTEKETTNNKQSTDG 300
QY 241 TQHNKVSSTPKITTAAPINRPSLPNSDTSKETSITVANKETVTEKETTNNKQSTDG 300
DB 241 TQHNKVSSTPKITTAAPINRPSLPNSDTSKETSITVANKETVTEKETTNNKQSTDG 300
QY 301 KEKTTSAKETOSIEKTSAKDLAPTSKVLAKPTPKAETTTGKALTTTPEPTTTPKEPAS 360
DB 301 KEKTTSAKETOSIEKTSAKDLAPTSKVLAKPTPKAETTTGKALTTTPEPTTTPKEPAS 360
QY 301 KEKTTSAKETOSIEKTSAKDLAPTSKVLAKPTPKAETTTGKALTTTPEPTTTPKEPAS 360
DB 301 KEKTTSAKETOSIEKTSAKDLAPTSKVLAKPTPKAETTTGKALTTTPEPTTTPKEPAS 360
QY 361 TTPKEPTTTPKSAPTTTPKEPAPTTTPKSAPTTTPKSAPTTTPKSAPTTTPKSAPTTTPK 420
DB 361 TTPKEPTTTPKSAPTTTPKEPAPTTTPKSAPTTTPKSAPTTTPKSAPTTTPKSAPTTTPK 420
QY 361 TTPKEPTTTPKSAPTTTPKEPAPTTTPKSAPTTTPKSAPTTTPKSAPTTTPKSAPTTTPK 420
DB 361 TTPKEPTTTPKSAPTTTPKEPAPTTTPKSAPTTTPKSAPTTTPKSAPTTTPKSAPTTTPK 420
QY 421 APTTTSAPTTTPKEPAPTTTPKSAPTTTPKSAPTTTPKSAPTTTPKSAPTTTPKSAPTTTPK 480
DB 421 APTTTSAPTTTPKEPAPTTTPKSAPTTTPKSAPTTTPKSAPTTTPKSAPTTTPKSAPTTTPK 480
QY 421 APTTTSAPTTTPKEPAPTTTPKSAPTTTPKSAPTTTPKSAPTTTPKSAPTTTPKSAPTTTPK 480
DB 421 APTTTSAPTTTPKEPAPTTTPKSAPTTTPKSAPTTTPKSAPTTTPKSAPTTTPKSAPTTTPK 480
QY 481 EPAPAPAPKAPAPTTTPKEPAPTTTPKSAPTTTPKSAPTTTPKSAPTTTPKSAPTTTPKSAPTTTPK 540
DB 481 EPAPAPAPKAPAPTTTPKEPAPTTTPKSAPTTTPKSAPTTTPKSAPTTTPKSAPTTTPKSAPTTTPK 540
QY 481 EPAPAPAPKAPAPTTTPKEPAPTTTPKSAPTTTPKSAPTTTPKSAPTTTPKSAPTTTPKSAPTTTPK 540
DB 481 EPAPAPAPKAPAPTTTPKEPAPTTTPKSAPTTTPKSAPTTTPKSAPTTTPKSAPTTTPKSAPTTTPK 540
QY 541 TTKSAPTTTPKESPTTTPKEPAPTTTPKSAPTTTPKSAPTTTPKSAPTTTPKSAPTTTPKSAPTTTPK 600
DB 541 TTKSAPTTTPKESPTTTPKEPAPTTTPKSAPTTTPKSAPTTTPKSAPTTTPKSAPTTTPKSAPTTTPK 600
QY 541 TTKSAPTTTPKESPTTTPKEPAPTTTPKSAPTTTPKSAPTTTPKSAPTTTPKSAPTTTPKSAPTTTPK 600
DB 541 TTKSAPTTTPKESPTTTPKEPAPTTTPKSAPTTTPKSAPTTTPKSAPTTTPKSAPTTTPKSAPTTTPK 600
QY 601 APTAKKEPAPTTTPKSAPTTTPKSAPTTTPKSAPTTTPKSAPTTTPKSAPTTTPKSAPTTTPKSAPTTTPK 660
DB 601 APTAKKEPAPTTTPKSAPTTTPKSAPTTTPKSAPTTTPKSAPTTTPKSAPTTTPKSAPTTTPKSAPTTTPK 660
QY 601 APTAKKEPAPTTTPKSAPTTTPKSAPTTTPKSAPTTTPKSAPTTTPKSAPTTTPKSAPTTTPKSAPTTTPK 660
DB 601 APTAKKEPAPTTTPKSAPTTTPKSAPTTTPKSAPTTTPKSAPTTTPKSAPTTTPKSAPTTTPKSAPTTTPK 660
QY 661 PEEBAPPTTTPKAAAPNTPEKAPAPTTTPKSAPTTTPKSAPTTTPKSAPTTTPKSAPTTTPKSAPTTTPK 720
DB 661 PEEBAPPTTTPKAAAPNTPEKAPAPTTTPKSAPTTTPKSAPTTTPKSAPTTTPKSAPTTTPKSAPTTTPK 720
QY 661 PEEBAPPTTTPKAAAPNTPEKAPAPTTTPKSAPTTTPKSAPTTTPKSAPTTTPKSAPTTTPKSAPTTTPK 720
DB 661 PEEBAPPTTTPKAAAPNTPEKAPAPTTTPKSAPTTTPKSAPTTTPKSAPTTTPKSAPTTTPKSAPTTTPK 720
QY 721 APTTTPKAPAPKELAPTTTKEPTSTSDKAPATTPKGTATTPKSAPTTTPKSAPTTTPKSAPTTTPKSAPTTTPK 780
DB 721 APTTTPKAPAPKELAPTTTKEPTSTSDKAPATTPKGTATTPKSAPTTTPKSAPTTTPKSAPTTTPKSAPTTTPK 780
QY 721 APTTTPKAPAPKELAPTTTKEPTSTSDKAPATTPKGTATTPKSAPTTTPKSAPTTTPKSAPTTTPKSAPTTTPK 780
DB 721 APTTTPKAPAPKELAPTTTKEPTSTSDKAPATTPKGTATTPKSAPTTTPKSAPTTTPKSAPTTTPKSAPTTTPK 780
QY 781 TAPTLKEBAPPTTTPKAPAPKELAPTTTKEPTSTSDKAPATTPKGTATTPKSAPTTTPKSAPTTTPKSAPTTTPK 840
DB 781 TAPTLKEBAPPTTTPKAPAPKELAPTTTKEPTSTSDKAPATTPKGTATTPKSAPTTTPKSAPTTTPKSAPTTTPK 840

```

```

DB 781 TAPTLKEBAPPTTTPKAPAPKELAPTTTKEPTSTSDKAPATTPKGTATTPKSAPTTTPKSAPTTTPKSAPTTTPK 840
QY 841 KPAPTTPETPTTSEVSTPTTKEPTTIKSPDESIPELSAETPALENSPEBPGVPT 900
DB 841 KPAPTTPETPTTSEVSTPTTKEPTTIKSPDESIPELSAETPALENSPEBPGVPT 900
QY 841 KPAPTTPETPTTSEVSTPTTKEPTTIKSPDESIPELSAETPALENSPEBPGVPT 900
DB 841 KPAPTTPETPTTSEVSTPTTKEPTTIKSPDESIPELSAETPALENSPEBPGVPT 900
QY 901 TKTPTAATPEWTTAKOTTERDRTTPETTAAPKMTKETATTEKTESKIRATTTQV 960
DB 901 TKTPTAATPEWTTAKOTTERDRTTPETTAAPKMTKETATTEKTESKIRATTTQV 960
QY 901 TKTPTAATPEWTTAKOTTERDRTTPETTAAPKMTKETATTEKTESKIRATTTQV 960
DB 901 TKTPTAATPEWTTAKOTTERDRTTPETTAAPKMTKETATTEKTESKIRATTTQV 960
QY 961 TSITTDOTPEKTTTLTKTTTLAPRVTTTKTTITTEIMNKEETAKPKDRATNSKATTPK 1020
DB 961 TSITTDOTPEKTTTLTKTTTLAPRVTTTKTTITTEIMNKEETAKPKDRATNSKATTPK 1020
QY 961 TSITTDOTPEKTTTLTKTTTLAPRVTTTKTTITTEIMNKEETAKPKDRATNSKATTPK 1020
DB 961 TSITTDOTPEKTTTLTKTTTLAPRVTTTKTTITTEIMNKEETAKPKDRATNSKATTPK 1020
QY 1021 POKTTPAKPKTSTTKKRTMPRVAKKPTTTPRKTSTMTPELNTSRIAEAMQTTTRPN 1080
DB 1021 POKTTPAKPKTSTTKKRTMPRVAKKPTTTPRKTSTMTPELNTSRIAEAMQTTTRPN 1080
QY 1021 POKTTPAKPKTSTTKKRTMPRVAKKPTTTPRKTSTMTPELNTSRIAEAMQTTTRPN 1080
DB 1021 POKTTPAKPKTSTTKKRTMPRVAKKPTTTPRKTSTMTPELNTSRIAEAMQTTTRPN 1080
QY 1081 QTPNSKLVENVPKSEDAAGAGETPHMLLRPHVMPVTPMDYLPVPNOGIIINPMLS 1140
DB 1081 QTPNSKLVENVPKSEDAAGAGETPHMLLRPHVMPVTPMDYLPVPNOGIIINPMLS 1140
QY 1081 QTPNSKLVENVPKSEDAAGAGETPHMLLRPHVMPVTPMDYLPVPNOGIIINPMLS 1140
DB 1081 QTPNSKLVENVPKSEDAAGAGETPHMLLRPHVMPVTPMDYLPVPNOGIIINPMLS 1140
QY 1141 DETNINCNGKPVDSGLTTLRNGTLVAFRGHYFWMLSPPSPSPARRITEVWGISPIDTVFT 1200
DB 1141 DETNINCNGKPVDSGLTTLRNGTLVAFRGHYFWMLSPPSPSPARRITEVWGISPIDTVFT 1200
QY 1141 DETNINCNGKPVDSGLTTLRNGTLVAFRGHYFWMLSPPSPSPARRITEVWGISPIDTVFT 1200
DB 1141 DETNINCNGKPVDSGLTTLRNGTLVAFRGHYFWMLSPPSPSPARRITEVWGISPIDTVFT 1200
QY 1201 RCNCEGKTFFPKDSQYRFTNDIDAGYPRPIFGFGGLTQYIYAALSTAKYKWPSSVY 1260
DB 1201 RCNCEGKTFFPKDSQYRFTNDIDAGYPRPIFGFGGLTQYIYAALSTAKYKWPSSVY 1260
QY 1201 RCNCEGKTFFPKDSQYRFTNDIDAGYPRPIFGFGGLTQYIYAALSTAKYKWPSSVY 1260
DB 1201 RCNCEGKTFFPKDSQYRFTNDIDAGYPRPIFGFGGLTQYIYAALSTAKYKWPSSVY 1260
QY 1261 FFKGGSIQYIYQOEYPQKCPGRPALANYPVYEMTOVRRRREARIGSQTHIRIQY 1320
DB 1261 FFKGGSIQYIYQOEYPQKCPGRPALANYPVYEMTOVRRRREARIGSQTHIRIQY 1320
QY 1261 FFKGGSIQYIYQOEYPQKCPGRPALANYPVYEMTOVRRRREARIGSQTHIRIQY 1320
DB 1261 FFKGGSIQYIYQOEYPQKCPGRPALANYPVYEMTOVRRRREARIGSQTHIRIQY 1320
QY 1321 SPARLAYQDKCVLHNEKVSILMRGLPNVVTSAISLPIRKPDSYDYAFSKQDYINIDV 1380
DB 1321 SPARLAYQDKCVLHNEKVSILMRGLPNVVTSAISLPIRKPDSYDYAFSKQDYINIDV 1380
QY 1381 PSRTARAIITRSQOTLSKWTNCP 1404
DB 1381 PSRTARAIITRSQOTLSKWTNCP 1404
QY 1381 PSRTARAIITRSQOTLSKWTNCP 1404
DB 1381 PSRTARAIITRSQOTLSKWTNCP 1404

```

RESULT 2
 US-10-124-557-62
 ; Sequence 62, Application US/10124557
 ; Publication No. US20020137894A1
 ; GENERAL INFORMATION:
 APPLICANT: Turner, Katherine
 Clark, Stephen C.
 Jacobs, Kenneth
 Hewick, Rodney M.
 Gesner, Thomas G.
 TITLE OF INVENTION: Megakaryocyte Stimulating Factors
 NUMBER OF SEQUENCES: 143
 CORRESPONDENCE ADDRESS:
 ADDRESSER: Genetics Institute, Inc.
 STREET: 87 Cambridgepark Drive
 CITY: Cambridge
 STATE: Massachusetts
 COUNTRY: U.S.A.
 ZIP: 02140
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/10/124,557
 FILING DATE: 16-Apr-2002
 CLASSIFICATION: <Unknown>
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/643,502

FILING DATE: 18-JAN-1991
 APPLICATION NUMBER: US 07/546,114
 FILING DATE: 29-JUN-1990
 APPLICATION NUMBER: US 07/457,196
 FILING DATE: 29-DEC-1989
 APPLICATION NUMBER: US 07/390,901
 FILING DATE: 08-AUG-1989
 ATTORNEY/AGENT INFORMATION:
 NAME: Gaert, Luann
 REGISTRATION NUMBER: 31,822
 REFERENCE/DOCKET NUMBER: GI 5190
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (617) 876-1170
 TELEFAX: (617) 876-5851
 INFORMATION FOR SEQ ID NO: 62:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1404 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 SEQUENCE DESCRIPTION: SEQ ID NO: 62:
 US-10-124-557-62

Query Match 100.0%; Score 7526; DB 14; Length 1404;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1404; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MAMKTPILYLLLSVAVVIOQVSSODLSCAGRCGEGSYSDATCNCDCYNCOHMECCPDF 60
 1 MAMKTPILYLLLSVAVVIOQVSSODLSCAGRCGEGSYSDATCNCDCYNCOHMECCPDF 60
 61 KRVCTABLSCGRCFESFERGRCDCDQCKKTDKCCPDYSEFCAEVHNTPSPSSKAP 120
 61 KRVCTABLSCGRCFESFERGRCDCDQCKKTDKCCPDYSEFCAEVHNTPSPSSKAP 120
 121 PRSGAQTIKSTTKRSBPNNKKTKXVISEEITTEHVSVENQESSSSSSSSSTI 180
 121 PRSGAQTIKSTTKRSBPNNKKTKXVISEEITTEHVSVENQESSSSSSSSSTI 180
 121 PRSGAQTIKSTTKRSBPNNKKTKXVISEEITTEHVSVENQESSSSSSSSSTI 180
 181 KISSKNSAANRELQKCLKYKDNKKKRTKKKPPKPVVDAAAGSLDNGFKATTDPTST 240
 181 KISSKNSAANRELQKCLKYKDNKKKRTKKKPPKPVVDAAAGSLDNGFKATTDPTST 240
 181 KISSKNSAANRELQKCLKYKDNKKKRTKKKPPKPVVDAAAGSLDNGFKATTDPTST 240
 241 TOHNVKSTSPKITTAKPINRPSLPNSDTSKETSILVNKETTVEKETTNNKOTSTG 300
 241 TOHNVKSTSPKITTAKPINRPSLPNSDTSKETSILVNKETTVEKETTNNKOTSTG 300
 241 TOHNVKSTSPKITTAKPINRPSLPNSDTSKETSILVNKETTVEKETTNNKOTSTG 300
 301 KEKTSKAKETOSIEKTSANDLAFTSKVLAKPTKAEITTTGPALETTTPKEPTTPKEPAS 360
 301 KEKTSKAKETOSIEKTSANDLAFTSKVLAKPTKAEITTTGPALETTTPKEPTTPKEPAS 360
 301 KEKTSKAKETOSIEKTSANDLAFTSKVLAKPTKAEITTTGPALETTTPKEPTTPKEPAS 360
 361 TTTPKEPTPTTIKAPPTTPKEPAATTTKSAATTPKEPAATTTKSAATTPKEPA 420
 361 TTTPKEPTPTTIKAPPTTPKEPAATTTKSAATTPKEPAATTTKSAATTPKEPA 420
 361 TTTPKEPTPTTIKAPPTTPKEPAATTTKSAATTPKEPAATTTKSAATTPKEPA 420
 421 APTTSAAPTTPKEPAATTPKCAPPTTPKEPAATTPKEPAATTPKEPAATTPKEPA 480
 421 APTTSAAPTTPKEPAATTPKCAPPTTPKEPAATTPKEPAATTPKEPAATTPKEPA 480
 421 APTTSAAPTTPKEPAATTPKCAPPTTPKEPAATTPKEPAATTPKEPAATTPKEPA 480
 481 EPAAPTAAPKKEPAATTPKEPAATTPKEPAATTPKEPASPTTPKEPAATTTKSA 540
 481 EPAAPTAAPKKEPAATTPKEPAATTPKEPAATTPKEPASPTTPKEPAATTTKSA 540
 481 EPAAPTAAPKKEPAATTPKEPAATTPKEPAATTPKEPASPTTPKEPAATTTKSA 540
 541 TTTSAPPTTPKEPSPTTPKEPAATTPKEPAATTPKKAAPTPKKAAPTPKKAAP 600
 541 TTTSAPPTTPKEPSPTTPKEPAATTPKEPAATTPKKAAPTPKKAAPTPKKAAP 600
 541 TTTSAPPTTPKEPSPTTPKEPAATTPKEPAATTPKKAAPTPKKAAPTPKKAAP 600
 601 APTAPKEPAATTPKEPAATTPKCLTPTPEKLAPTTPEKAPATTPPEELATTTSEBPPT 660
 601 APTAPKEPAATTPKEPAATTPKCLTPTPEKLAPTTPEKAPATTPPEELATTTSEBPPT 660
 601 APTAPKEPAATTPKEPAATTPKCLTPTPEKLAPTTPEKAPATTPPEELATTTSEBPPT 660
 661 PEEPAATTPKAAAPNTPEKEPAATTPKEPAATTPKEPAATTPKETAATTPKGTAPTTLKEP 720
 661 PEEPAATTPKAAAPNTPEKEPAATTPKEPAATTPKEPAATTPKETAATTPKGTAPTTLKEP 720
 661 PEEPAATTPKAAAPNTPEKEPAATTPKEPAATTPKEPAATTPKETAATTPKGTAPTTLKEP 720

DB 661 PEEPAATTPKAAAPNTPEKEPAATTPKEPAATTPKEPAATTPKETAATTPKGTAPTTLKEP 720
 QY 721 APTPKKAPKELAPPTTKEPTSTSDKAPPTTPKGTATTPKEPAATTPKKEPAATTPG 780
 DB 721 APTPKKAPKELAPPTTKEPTSTSDKAPPTTPKGTATTPKEPAATTPKKEPAATTPG 780
 QY 781 TAPTTLKEPAATTPKKEPAATTPKGTSTSDKAPPTTPKETAATTPKEPAATTPK 840
 DB 781 TAPTTLKEPAATTPKKEPAATTPKGTSTSDKAPPTTPKETAATTPKEPAATTPK 840
 QY 841 KAPPTPEPTPEPTSEVSTPTTKEPTTIHKS PDSPELSAEPKALENSKEGVP 900
 DB 841 KAPPTPEPTPEPTSEVSTPTTKEPTTIHKS PDSPELSAEPKALENSKEGVP 900
 QY 901 TKTPAATKEMTTAKOKTTERDLRTPPTTAAAPMTKETAATTEKTESKITATTTQV 960
 DB 901 TKTPAATKEMTTAKOKTTERDLRTPPTTAAAPMTKETAATTEKTESKITATTTQV 960
 QY 961 TSTTODTTPFKITTLKTTTLAPKVTTKKITTTEINMKPESTAKPKDRAATNSKATTPK 1020
 DB 961 TSTTODTTPFKITTLKTTTLAPKVTTKKITTTEINMKPESTAKPKDRAATNSKATTPK 1020
 QY 1021 POKPTAPKPKPTSTKPKTMPVRKPKTTPRCKMTSTMPBELNPTSRIAEMLOTTTRBN 1080
 DB 1021 POKPTAPKPKPTSTKPKTMPVRKPKTTPRCKMTSTMPBELNPTSRIAEMLOTTTRBN 1080
 QY 1081 QTPNSKLVVNPKSEBAGABEETPMMLRPVFMVEVTPDMDYLRPVNOCIIINPMIS 1140
 DB 1081 QTPNSKLVVNPKSEBAGABEETPMMLRPVFMVEVTPDMDYLRPVNOCIIINPMIS 1140
 QY 1141 DETNLCNGKPVNGLATLRNGTLVAFRGHYFMMLSPPSPARRITEVNGISPIDTVP 1200
 DB 1141 DETNLCNGKPVNGLATLRNGTLVAFRGHYFMMLSPPSPARRITEVNGISPIDTVP 1200
 QY 1201 RCNCEGTEFFKDSQYWRFTNDIKDAGYKPIFKGFGLTGQIVAAALSTAKYKMBESY 1260
 DB 1201 RCNCEGTEFFKDSQYWRFTNDIKDAGYKPIFKGFGLTGQIVAAALSTAKYKMBESY 1260
 QY 1261 FRKRGSIQOYLYKQEPVQKCGRRPALNVPYVGENTQVRRRRFEPALIPSOHTTIRIQ 1320
 DB 1261 FRKRGSIQOYLYKQEPVQKCGRRPALNVPYVGENTQVRRRRFEPALIPSOHTTIRIQ 1320
 QY 1321 SPARLAADKGVLAHNEVKVSIIMRGLPNVTSAISLPNIRKPDGYDYAFSDQOYINIDV 1380
 DB 1321 SPARLAADKGVLAHNEVKVSIIMRGLPNVTSAISLPNIRKPDGYDYAFSDQOYINIDV 1380
 QY 1381 PRTARAITTRSGQTLISKVYNCP 1404
 DB 1381 PRTARAITTRSGQTLISKVYNCP 1404

RESULT 3
 US-10-124-557-40
 Sequence 40, Application US/10124557
 Publication No. US20020137894A1
 GENERAL INFORMATION:
 APPLICANT: Turner, Katherine
 Clark, Stephen C.
 Jacobs, Kenneth
 Hewick, Rodney M.
 Geener, Thomas G.
 TITLE OF INVENTION: Megakaryocyte Stimulating Factors
 NUMBER OF SEQUENCES: 143
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Genetics Institute, Inc.
 STREET: 87 Cambridgepark Drive
 CITY: Cambridge
 STATE: Massachusetts
 COUNTRY: U.S.A.
 ZIP: 02140
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/124,557
FILING DATE: 16-Apr-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JUN-1991
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Caer, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 876-1170
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 40:
SEQUENCE CHARACTERISTICS:
LENGTH: 1361 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 40:
US-10-124-557-40

Query Match 96.9%; Score 7291.5; DB 14; Length 1361;
Best Local Similarity 96.9%; Pred. No. 0;
Matches 1361; Conservative 0; Mismatches 0; Indels 43; Gaps 1;

QY 1 MAMKTLPIYLILLISLVFVIQOVSSQDLSSCAGRCGEYSRDATCNDYNCQHYMECCPDF 60
DB 1 MAMKTLPIYLILLISLVFVIQOVSSQDLSSCAGRCGEYSRDATCNDYNCQHYMECCPDF 60
QY 61 KRVTALISCKGRCESEFERGECDDAQCKKYDKCCPDYSEFCALVHNPTSPSSSKAP 120
DB 61 KRVTALISCKGRCESEFERGECDDAQCKKYDKCCPDYSEFCALVHNPTSPSSSKAP 120
QY 61 KRVTALISCKGRCESEFERGECDDAQCKKYDKCCPDYSEFCALVHNPTSPSSSKAP 120
DB 61 KRVTALISCKGRCESEFERGECDDAQCKKYDKCCPDYSEFCALVHNPTSPSSSKAP 120
QY 121 PPSGASQITKSTKSPKPKKKTKKYESEETIEHSVSENOSSSSSSSSSSSSSSSSSSSS 180
DB 121 PPSGASQITKSTKSPKPKKKTKKYESEETIEHSVSENOSSSSSSSSSSSSSSSSSSSS 180
QY 121 PPSGASQITKSTKSPKPKKKTKKYESEETIEHSVSENOSSSSSSSSSSSSSSSSSSSS 180
DB 121 PPSGASQITKSTKSPKPKKKTKKYESEETIEHSVSENOSSSSSSSSSSSSSSSSSSSS 180
QY 181 KIKSSKNSANRELQKLLKVDNKKNKRTKKKTPKPPVVDAGSGLDNDGPKVTTPTST 240
DB 181 KIKSSKNSANRELQKLLKVDNKKNKRTKKKTPKPPVVDAGSGLDNDGPKVTTPTST 240
QY 157 -----VDNKKNKRTKKKTPKPPVVDAGSGLDNDGPKVTTPTST 197
DB 157 -----VDNKKNKRTKKKTPKPPVVDAGSGLDNDGPKVTTPTST 197
QY 241 TQHNKVSPTKITAKPINRPSLPNSDTSKETSIVNKKETVETKETTNNKQTSIDG 300
DB 241 TQHNKVSPTKITAKPINRPSLPNSDTSKETSIVNKKETVETKETTNNKQTSIDG 300
QY 198 TQHNKVSPTKITAKPINRPSLPNSDTSKETSIVNKKETVETKETTNNKQTSIDG 257
DB 198 TQHNKVSPTKITAKPINRPSLPNSDTSKETSIVNKKETVETKETTNNKQTSIDG 257
QY 301 KEKTSKETSISKTSKDLAPTSKYLAKTPKAKETTKGAPALTTKPEPTTPPKKESAS 360
DB 301 KEKTSKETSISKTSKDLAPTSKYLAKTPKAKETTKGAPALTTKPEPTTPPKKESAS 360
QY 258 KEKTSKETSISKTSKDLAPTSKYLAKTPKAKETTKGAPALTTKPEPTTPPKKESAS 317
DB 258 KEKTSKETSISKTSKDLAPTSKYLAKTPKAKETTKGAPALTTKPEPTTPPKKESAS 317
QY 361 TTPKPEPTTPPKKESASPTTKKAPALTTKPEPTTPPKKAPALTTKPEPTTPPKKAP 420
DB 361 TTPKPEPTTPPKKESASPTTKKAPALTTKPEPTTPPKKAPALTTKPEPTTPPKKAP 420
QY 318 TTPKPEPTTPPKKESASPTTKKAPALTTKPEPTTPPKKAPALTTKPEPTTPPKKAP 377
DB 318 TTPKPEPTTPPKKESASPTTKKAPALTTKPEPTTPPKKAPALTTKPEPTTPPKKAP 377
QY 421 APPTTKSAPTPPKKAPALTTKPEPTTPPKKAPALTTKPEPTTPPKKAPALTTK 480
DB 421 APPTTKSAPTPPKKAPALTTKPEPTTPPKKAPALTTKPEPTTPPKKAPALTTK 480
QY 378 APPTTKSAPTPPKKAPALTTKPEPTTPPKKAPALTTKPEPTTPPKKAPALTTK 437
DB 378 APPTTKSAPTPPKKAPALTTKPEPTTPPKKAPALTTKPEPTTPPKKAPALTTK 437
QY 481 EPAPTAAPKAPALTTKPEPTTPPKKAPALTTKPEPTTPPKKAPALTTKSAPTTK 540
DB 481 EPAPTAAPKAPALTTKPEPTTPPKKAPALTTKPEPTTPPKKAPALTTKSAPTTK 540
QY 438 EPAPTAAPKAPALTTKPEPTTPPKKAPALTTKPEPTTPPKKAPALTTKSAPTTK 497
DB 438 EPAPTAAPKAPALTTKPEPTTPPKKAPALTTKPEPTTPPKKAPALTTKSAPTTK 497
QY 541 TTKAPALTTKPEPTTPPKKAPALTTKPEPTTPPKKAPALTTKPEPTTPPKKAP 600
DB 541 TTKAPALTTKPEPTTPPKKAPALTTKPEPTTPPKKAPALTTKPEPTTPPKKAP 600

DB 498 TTKAPALTTKPEPTTPPKKAPALTTKPEPTTPPKKAPALTTKPEPTTPPKKAP 557
QY 601 APAPKAPALTTKPEPTTPPKKAPALTTKPEPTTPPKKAPALTTKPEPTTPPKKAP 660
DB 558 APAPKAPALTTKPEPTTPPKKAPALTTKPEPTTPPKKAPALTTKPEPTTPPKKAP 617
QY 661 PEPAPALTTKPEPTTPPKKAPALTTKPEPTTPPKKAPALTTKPEPTTPPKKAP 720
DB 618 PEPAPALTTKPEPTTPPKKAPALTTKPEPTTPPKKAPALTTKPEPTTPPKKAP 677
QY 721 APPTTKKAPALTTKPEPTTPPKKAPALTTKPEPTTPPKKAPALTTKPEPTTPPKK 780
DB 678 APPTTKKAPALTTKPEPTTPPKKAPALTTKPEPTTPPKKAPALTTKPEPTTPPKK 737
QY 781 TAPPTTKKAPALTTKPEPTTPPKKAPALTTKPEPTTPPKKAPALTTKPEPTTPPKK 840
DB 738 TAPPTTKKAPALTTKPEPTTPPKKAPALTTKPEPTTPPKKAPALTTKPEPTTPPKK 797
QY 841 KPAPTEPTTPPKKAPALTTKPEPTTPPKKAPALTTKPEPTTPPKKAPALTTKPEPTTP 900
DB 798 KPAPTEPTTPPKKAPALTTKPEPTTPPKKAPALTTKPEPTTPPKKAPALTTKPEPTTP 857
QY 901 TKTAPATKPEPTTPPKKAPALTTKPEPTTPPKKAPALTTKPEPTTPPKKAPALTT 960
DB 858 TKTAPATKPEPTTPPKKAPALTTKPEPTTPPKKAPALTTKPEPTTPPKKAPALTT 917
QY 961 TSTTTQDTPPKKAPALTTKPEPTTPPKKAPALTTKPEPTTPPKKAPALTTKPEPTTP 1020
DB 918 TSTTTQDTPPKKAPALTTKPEPTTPPKKAPALTTKPEPTTPPKKAPALTTKPEPTTP 977
QY 1021 POKPTKAPKPEPTTPPKKAPALTTKPEPTTPPKKAPALTTKPEPTTPPKKAPALTT 1080
DB 978 POKPTKAPKPEPTTPPKKAPALTTKPEPTTPPKKAPALTTKPEPTTPPKKAPALTT 1037
QY 1081 QTPNSKLVENPKSEDAGAGETPMILRPHVMEVETDMDYLPRVPMGIIINPMLS 1140
DB 1038 QTPNSKLVENPKSEDAGAGETPMILRPHVMEVETDMDYLPRVPMGIIINPMLS 1097
QY 1141 DETNINCKGKPVGDLITLRNGTLVAFRGHYFMMLSPESPSPARITTEWGIPIIDIVFT 1200
DB 1098 DETNINCKGKPVGDLITLRNGTLVAFRGHYFMMLSPESPSPARITTEWGIPIIDIVFT 1157
QY 1201 RNCCEGKTFPFKDSQVRFNDIDAGYPRIFKFGGLGQIYVAAALSTAKYKMPESVY 1260
DB 1158 RNCCEGKTFPFKDSQVRFNDIDAGYPRIFKFGGLGQIYVAAALSTAKYKMPESVY 1217
QY 1261 FFKRGSIIQYIYVQEPVQKCPGRPALNYPVYGEMLQVRRRRERAIIGSQTHTIRIQY 1320
DB 1218 FFKRGSIIQYIYVQEPVQKCPGRPALNYPVYGEMLQVRRRRERAIIGSQTHTIRIQY 1277
QY 1321 SPARLAVQDKGVLANEVKVSILMRGLPNVYTSALSLNIRKPDGYDYAASKOYVINDV 1380
DB 1278 SPARLAVQDKGVLANEVKVSILMRGLPNVYTSALSLNIRKPDGYDYAASKOYVINDV 1337
QY 1381 PSRTARAITRSQOTLSKWNCP 1404
DB 1338 PSRTARAITRSQOTLSKWNCP 1361

RESULT 4
US-10-124-557-52
Sequence 52, Application US/10124557
Publication No. US20020137894A1
GENERAL INFORMATION:
APPLICANT: Turner, Katherine
Clark, Stephen C.
Jacobs, Kenneth
Hewick, Rodney M.
Gesner, Thomas G.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.


```

/ STREET: 87 Cambridgepark Drive
/ CITY: Cambridge
/ STATE: Massachusetts
/ COUNTRY: U.S.A.
/ ZIP: 02140
/
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patentin Releasee #1.0, Version #1.25
/
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/10/124,557
/ FILING DATE: 16-Apr-2002
/ CLASSIFICATION: <Unknown>
/
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 07/643,502
/ FILING DATE: 18-JAN-1991
/ APPLICATION NUMBER: US 07/546,114
/ FILING DATE: 29-JUN-1990
/ APPLICATION NUMBER: US 07/457,196
/ FILING DATE: 29-DEC-1989
/ APPLICATION NUMBER: US 07/390,901
/
/ FILING DATE: 08-AUG-1989
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Coert, Luann
/ REGISTRATION NUMBER: 31,822
/ REFERENCE/DOCKET NUMBER: GI 5190
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (617)876-1170
/ TELEFAX: (617)876-5851
/
/ INFORMATION FOR SEQ ID NO: 52:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 1363 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/ SEQUENCE DESCRIPTION: SEQ ID NO: 52:
US-10-124-557-52

Query Match      96.3%; Score 7245.5; DB 14; Length 1363;
Best Local Similarity 97.1%; Pred. No. 0;
Matches 1363; Conservative 0; Mismatches 0; Indels 41; Gaps 1;

QY 1 MAMKTLPIYLILLVFIQVVSQDLSGACRGCEGYSRDATCNDVNCQHYMECCPFP 60
DB 1 MAMKTLPIYLILLVFIQVVSQDLSGACRGCEGYSRDATCNDVNCQHYMECCPFP 25
QY 61 KKVCTAELSGCRCEPESFERGREGCDCAQCKYDKCCPDYESFCAEVHNPTSPSSKAP 120
DB 61 KKVCTAELSGCRCEPESFERGREGCDCAQCKYDKCCPDYESFCAEVHNPTSPSSKAP 79
QY 26 -----ELSGCRCEPESFERGREGCDCAQCKYDKCCPDYESFCAEVHNPTSPSSKAP 79
DB 26 -----ELSGCRCEPESFERGREGCDCAQCKYDKCCPDYESFCAEVHNPTSPSSKAP 79
QY 121 PPSGASQITKSTTKSGSPKPPNKKTKTKVIESEBITTEHSVSNQESSSSSSSSSTIW 180
DB 121 PPSGASQITKSTTKSGSPKPPNKKTKTKVIESEBITTEHSVSNQESSSSSSSSSTIW 139
QY 80 PPSGASQITKSTTKSGSPKPPNKKTKTKVIESEBITTEHSVSNQESSSSSSSSSTIW 139
DB 80 PPSGASQITKSTTKSGSPKPPNKKTKTKVIESEBITTEHSVSNQESSSSSSSSSTIW 139
QY 181 KIKSGNSANRELQKKLVKONKKRRTKKKTPRPVVUDEAGSGLDNDKVTTPDST 240
DB 181 KIKSGNSANRELQKKLVKONKKRRTKKKTPRPVVUDEAGSGLDNDKVTTPDST 199
QY 140 KIKSSNSANRELQKKLVKONKKRRTKKKTPRPVVUDEAGSGLDNDKVTTPDST 199
DB 140 KIKSSNSANRELQKKLVKONKKRRTKKKTPRPVVUDEAGSGLDNDKVTTPDST 199
QY 241 TOHNKVSPTKITTAKPINSPLPNSDTSKETSITVAKETTVETKETTTNKKQSTDG 300
DB 241 TOHNKVSPTKITTAKPINSPLPNSDTSKETSITVAKETTVETKETTTNKKQSTDG 259
QY 200 TOHNKVSPTKITTAKPINSPLPNSDTSKETSITVAKETTVETKETTTNKKQSTDG 259
DB 200 TOHNKVSPTKITTAKPINSPLPNSDTSKETSITVAKETTVETKETTTNKKQSTDG 259
QY 301 KERTTSANETOSIEKTSADLAPTSKVLAKPTPKAETTTGPAITTPKEPTTPKEPAS 360
DB 301 KERTTSANETOSIEKTSADLAPTSKVLAKPTPKAETTTGPAITTPKEPTTPKEPAS 319
QY 260 KERTTSANETOSIEKTSADLAPTSKVLAKPTPKAETTTGPAITTPKEPTTPKEPAS 319
DB 260 KERTTSANETOSIEKTSADLAPTSKVLAKPTPKAETTTGPAITTPKEPTTPKEPAS 319
QY 361 TTPKEPTTPKISAPTTKEPAPTTTKSAPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEP 420
DB 361 TTPKEPTTPKISAPTTKEPAPTTTKSAPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEP 420
QY 320 TTPKEPTTPKISAPTTKEPAPTTTKSAPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEP 379
DB 320 TTPKEPTTPKISAPTTKEPAPTTTKSAPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEP 379
QY 421 APTTTKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTPK 480
DB 421 APTTTKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTPK 480

```

```

DB 360 APTTTKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTPK 439
QY 461 EPAPTAAPKKAAPTTPKEPAPTTTKEPAPTTTKEPSPTTPKEPAPTTTKSAPTTTKEPAPT 540
DB 461 EPAPTAAPKKAAPTTPKEPAPTTTKEPAPTTTKEPSPTTPKEPAPTTTKSAPTTTKEPAPT 499
QY 541 TTKSAPTTPKEPSPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPA 600
DB 541 TTKSAPTTPKEPSPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPA 559
QY 500 TTKSAPTTPKEPSPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPA 559
DB 500 TTKSAPTTPKEPSPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPA 559
QY 601 APTAKKEPAPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 660
DB 601 APTAKKEPAPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 619
QY 560 APTAKKEPAPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 619
DB 560 APTAKKEPAPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 619
QY 661 PEEPAAPTTPKAAAPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPA 720
DB 661 PEEPAAPTTPKAAAPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPA 679
QY 620 PEEPAAPTTPKAAAPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPA 679
DB 620 PEEPAAPTTPKAAAPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPA 679
QY 721 APTTPKEPAPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 780
DB 721 APTTPKEPAPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 739
QY 660 APTTPKEPAPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 739
DB 660 APTTPKEPAPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 739
QY 781 TAPTTKEPAPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 840
DB 781 TAPTTKEPAPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 799
QY 740 TAPTTKEPAPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 799
DB 740 TAPTTKEPAPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 799
QY 841 KPAPTTPPTPTTSEVSTPTTKKEPTTIHKSPEDESTPELSAETPKALENSPKPGVPT 900
DB 841 KPAPTTPPTPTTSEVSTPTTKKEPTTIHKSPEDESTPELSAETPKALENSPKPGVPT 859
QY 800 KPAPTTPPTPTTSEVSTPTTKKEPTTIHKSPEDESTPELSAETPKALENSPKPGVPT 859
DB 800 KPAPTTPPTPTTSEVSTPTTKKEPTTIHKSPEDESTPELSAETPKALENSPKPGVPT 859
QY 901 TKTPAATKPEMTTAKDITTEBDLTTPETTTAAPKMTKEATTTTEKTESKITATTTQV 960
DB 901 TKTPAATKPEMTTAKDITTEBDLTTPETTTAAPKMTKEATTTTEKTESKITATTTQV 919
QY 860 TKTPAATKPEMTTAKDITTEBDLTTPETTTAAPKMTKEATTTTEKTESKITATTTQV 919
DB 860 TKTPAATKPEMTTAKDITTEBDLTTPETTTAAPKMTKEATTTTEKTESKITATTTQV 919
QY 961 TSTTQDTPPTTKITLTKTTTLAPKVTYTKKTIITTEINMKDEETAPKPDATNSKATTPK 1020
DB 961 TSTTQDTPPTTKITLTKTTTLAPKVTYTKKTIITTEINMKDEETAPKPDATNSKATTPK 979
QY 920 TSTTQDTPPTTKITLTKTTTLAPKVTYTKKTIITTEINMKDEETAPKPDATNSKATTPK 979
DB 920 TSTTQDTPPTTKITLTKTTTLAPKVTYTKKTIITTEINMKDEETAPKPDATNSKATTPK 979
QY 1021 POKPTKAPKPTSTYKKEPTMFRVRKPKTTPPRKMTSTMPBELNPTSRIAEAMLQTTTRPN 1080
DB 1021 POKPTKAPKPTSTYKKEPTMFRVRKPKTTPPRKMTSTMPBELNPTSRIAEAMLQTTTRPN 1039
QY 980 POKPTKAPKPTSTYKKEPTMFRVRKPKTTPPRKMTSTMPBELNPTSRIAEAMLQTTTRPN 1039
DB 980 POKPTKAPKPTSTYKKEPTMFRVRKPKTTPPRKMTSTMPBELNPTSRIAEAMLQTTTRPN 1039
QY 1081 QTPNSKLVENVPKSDDAGAGETPHMLLRHVMPPEATTPMDVLPAPNGGIIINPMLS 1140
DB 1081 QTPNSKLVENVPKSDDAGAGETPHMLLRHVMPPEATTPMDVLPAPNGGIIINPMLS 1099
QY 1040 QTPNSKLVENVPKSDDAGAGETPHMLLRHVMPPEATTPMDVLPAPNGGIIINPMLS 1099
DB 1040 QTPNSKLVENVPKSDDAGAGETPHMLLRHVMPPEATTPMDVLPAPNGGIIINPMLS 1099
QY 1141 DETNINCKGAPVDGLITTLANGTLVAFRGHYFMMLSPSPSPARITTEVWGIISPIDTVFT 1200
DB 1141 DETNINCKGAPVDGLITTLANGTLVAFRGHYFMMLSPSPSPARITTEVWGIISPIDTVFT 1159
QY 1100 DETNINCKGAPVDGLITTLANGTLVAFRGHYFMMLSPSPSPARITTEVWGIISPIDTVFT 1159
DB 1100 DETNINCKGAPVDGLITTLANGTLVAFRGHYFMMLSPSPSPARITTEVWGIISPIDTVFT 1159
QY 1201 RNCCEGKTFPFDKSOYWFPTNDIKDAGYKPKIFKPGGLTGOIYVAALSTANYKWPESVY 1260
DB 1201 RNCCEGKTFPFDKSOYWFPTNDIKDAGYKPKIFKPGGLTGOIYVAALSTANYKWPESVY 1219
QY 1160 RNCCEGKTFPFDKSOYWFPTNDIKDAGYKPKIFKPGGLTGOIYVAALSTANYKWPESVY 1219
DB 1160 RNCCEGKTFPFDKSOYWFPTNDIKDAGYKPKIFKPGGLTGOIYVAALSTANYKWPESVY 1219
QY 1261 FFKGGGSIQOYIYKOEPOKQGRRRPALNVYVGEQVRRRRPBRALGSGQTHIRIQY 1320
DB 1261 FFKGGGSIQOYIYKOEPOKQGRRRPALNVYVGEQVRRRRPBRALGSGQTHIRIQY 1279
QY 1220 FFKGGGSIQOYIYKOEPOKQGRRRPALNVYVGEQVRRRRPBRALGSGQTHIRIQY 1279
DB 1220 FFKGGGSIQOYIYKOEPOKQGRRRPALNVYVGEQVRRRRPBRALGSGQTHIRIQY 1279
QY 1321 SPARLAYODKGVNLNEVVSILMRGLPNVYVSAISLPIRRKPDGDYVAFKQDYNNIDV 1380
DB 1321 SPARLAYODKGVNLNEVVSILMRGLPNVYVSAISLPIRRKPDGDYVAFKQDYNNIDV 1339
QY 1280 SPARLAYODKGVNLNEVVSILMRGLPNVYVSAISLPIRRKPDGDYVAFKQDYNNIDV 1339
DB 1280 SPARLAYODKGVNLNEVVSILMRGLPNVYVSAISLPIRRKPDGDYVAFKQDYNNIDV 1339
QY 1381 PSRTARAITTSGQTLKVMYVNC 1404
DB 1381 PSRTARAITTSGQTLKVMYVNC 1363
QY 1340 PSRTARAITTSGQTLKVMYVNC 1363
DB 1340 PSRTARAITTSGQTLKVMYVNC 1363

```

RESULT 5
 US-10-124-557-48
 ; Sequence 48, Application US/10124557
 ; Publication No. US20020137894A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Turner, Katherine

RESULT 6
US-10-124-557-46
Sequence 46, Application US/10124557
Publication No. US20020137894A1
GENERAL INFORMATION:
APPLICANT: Turner, Katherine
Clark, Stephen C.
Jacob, Kenneth
Hewick, Rodney M.
Gessner, Thomas G.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/124,557
FILING DATE: 16-Apr-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cbeert, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 46:
SEQUENCE CHARACTERISTICS:
LENGTH: 1320 amino acids
TYPE: amino acid
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 46:
US-10-124-557-46

Query Match 93.2%; Score 7011; DB 14; Length 1320;
Best Local Similarity 94.0%; Pred. No. 0;
Matches 1340; Conservative 0; Mismatches 0; Indels 84; Gaps 2;

QY 1 MAMKTIPIYLLLLSVFVIQVSSODLSACRGGYSHDATCNCDYNCQHYMECCPPF 60
DB 1 MAMKTIPIYLLLLSVFVIQVSSO----- 25
QY 61 KRVCTNELSGKRCFSPERGRGECDDACQCKYDKCCPDYSEFCAEVHNFTSPSSKAP 120
DB 26 -----ELSGKRCFSPERGRGECDDACQCKYDKCCPDYSEFCAEVHNFTSPSSKAP 79
QY 121 PPSGAGQTTKSTTKRSPKPKNKTKKVKVSESEITEHSEVSENGSSSSSSSSSSSTM 180
DB 80 PPSGAGQTTKSTTKRSPKPKNKTKKVKVSESEITE----- 115
QY 181 KIKSKIKSAAARELQKKLKVKNKKKRTKKKPTPKPVVDAGSGLDNGDFKVTTPDTST 240
DB 115 KIKSKIKSAAARELQKKLKVKNKKKRTKKKPTPKPVVDAGSGLDNGDFKVTTPDTST 240

DB 116 -----VKDNKKRRTKKKPTPKPVVDAGSGLDNGDFKVTTPDTST 156
QY 241 TQHNKYSTPKITTAKEINPRPSLPNSDTSKETSITVKKETVETKETTNNKQSTOG 300
DB 157 TQHNKYSTPKITTAKEINPRPSLPNSDTSKETSITVKKETVETKETTNNKQSTOG 216
QY 301 KEKTSKAKETQSIKTSKADLAPTSKVLAKPTPKAETTTKGPALTPKEPTPTTPEKPS 360
DB 217 KEKTSKAKETQSIKTSKADLAPTSKVLAKPTPKAETTTKGPALTPKEPTPTTPEKPS 276
QY 361 TTPKEPTPTTISAPTTKEPAPTTTSKAPTTTKEPAPTTKEPAPTTKEPAPTTKEP 420
DB 277 TTPKEPTPTTISAPTTKEPAPTTTSKAPTTTKEPAPTTKEPAPTTKEPAPTTKEP 336
QY 421 APTTTSKAPTTKEPAPTTTPKPAPTTTPKEPAPTTTPKEPTPTTTPKEPAPTTKEP 480
DB 337 APTTTSKAPTTKEPAPTTTPKPAPTTTPKEPAPTTTPKEPTPTTTPKEPAPTTKEP 396
QY 481 EPAPTAPKKAPATTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPA 540
DB 397 EPAPTAPKKAPATTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPA 456
QY 541 TTKSAPTTKEPSPTTKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPA 600
DB 457 TTKSAPTTKEPSPTTKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPA 516
QY 601 APTAPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPA 660
DB 517 APTAPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPA 576
QY 661 PEPAPTTPKAAAPNTPKAPATTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPA 720
DB 577 PEPAPTTPKAAAPNTPKAPATTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPA 636
QY 721 APTTPKPAKELAPTTTKEPTSTSDKAPATTTPKGAATTTPKEPAPTTTPKEPA 780
DB 637 APTTPKPAKELAPTTTKEPTSTSDKAPATTTPKGAATTTPKEPAPTTTPKEPA 696
QY 781 TAPTTKEPAPTTTPKPAKELAPTTTKEPTSTSDKAPATTTPKGAATTTPKEPA 840
DB 697 TAPTTKEPAPTTTPKPAKELAPTTTKEPTSTSDKAPATTTPKGAATTTPKEPA 756
QY 841 KAPPTPEPTPTTSTSVSTTTTKEPTTHKSPDESTPLSAPTPKALENSKEGVP 900
DB 757 KAPPTPEPTPTTSTSVSTTTTKEPTTHKSPDESTPLSAPTPKALENSKEGVP 816
QY 901 TKTPATKEMTTAKDKTERDLRTTPETTAAPTKEATTTTEKTESKITATTTQV 960
DB 817 TKTPATKEMTTAKDKTERDLRTTPETTAAPTKEATTTTEKTESKITATTTQV 876
QY 961 TSTTTODTTPKITTILKTTTLPKVTTKKITTTEIANKPBEITAKKORATNSKATTPK 1020
DB 877 TSTTTODTTPKITTILKTTTLPKVTTKKITTTEIANKPBEITAKKORATNSKATTPK 936
QY 1021 POKPTAPKKPTSTKKPKTPVRKPTTPTRKMTSTMPBLNPTSRIEAMQTTTRN 1080
DB 937 POKPTAPKKPTSTKKPKTPVRKPTTPTRKMTSTMPBLNPTSRIEAMQTTTRN 996
QY 1081 QTPNSKLVENPKSEBAGABETPHMLRPVFMPEVTPDMVDYLPRVNOGIIINPMS 1140
DB 997 QTPNSKLVENPKSEBAGABETPHMLRPVFMPEVTPDMVDYLPRVNOGIIINPMS 1056
QY 1141 DETNINCGRVUGLTLNGLTLVATRGHYFMMLSPSPSPSPARRITTEVWGIFSPIDVT 1200
DB 1057 DETNINCGRVUGLTLNGLTLVATRGHYFMMLSPSPSPSPARRITTEVWGIFSPIDVT 1116
QY 1201 RCNCGKTFEFDOSQVFRFTNDIXDAGVKKPIFKGCGGLTGQIVAAALSTAKTKMPE 1260
DB 1117 RCNCGKTFEFDOSQVFRFTNDIXDAGVKKPIFKGCGGLTGQIVAAALSTAKTKMPE 1176
QY 1261 PFKRGSIQOYIYKQBPVQKCGRRPALNYPVYGEMTVQRRRRFERAIGSQTHTRIOY 1320
DB 1177 PFKRGSIQOYIYKQBPVQKCGRRPALNYPVYGEMTVQRRRRFERAIGSQTHTRIOY 1236

1321 SPARLAYODKGVLEHNEVAVSILMRGLPMVVTSAISLPMIRKPDGYDYAFASDQYNNIDV 1380
1237 SPARLAYODKGVLEHNEVAVSILMRGLPMVVTSAISLPMIRKPDGYDYAFASDQYNNIDV 1296
1381 PSRTARAITTRSQGLTSKVMYNCP 1404
1297 PSRTARAITTRSQGLTSKVMYNCP 1320

RESULT 7
US-10-124-557-60
Sequence 60, Application US/10124557
Publication No. US20020137894A1
GENERAL INFORMATION:
APPLICANT: Turner, Katherine
Clark, Stephen C.
Jacobs, Kenneth
Hewick, Rodney M.
Gesner, Thomas G.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/124,557
FILING DATE: 16-Apr-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Caertr, Luann
REGISTRATION NUMBER: 31, 822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 876-1170
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 60:
SEQUENCE CHARACTERISTICS:
LENGTH: 1320 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 60:
US-10-124-557-60

Query Match 93.2%; Score 7011; DB 14; Length 1320;
Best Local Similarity 94.0%; Pred. No. 0;
Matches 1320; Conservative 0; Mismatches 0; Indels 84; Gaps 2;

QY 1 MAMWTLPIYLLLSVFVIOVSSODLSSCAGRCGEGSRATNCQVNCQHWECDF 60
DB 1 MAMWTLPIYLLLSVFVIOVSSODLSSCAGRCGEGSRATNCQVNCQHWECDF 25
QY 61 KRVCTAELSCGRCFESFERGRECDCAQCCKYDKCCPDYESFCAEVNPTSPSSKKAP 120

DB 26 -----ELSCGRCFESFERGRECDCAQCCKYDKCCPDYESFCAEVNPTSPSSKKAP 79
QY 121 PPSGASQITKSTKRSPPKPKKTKKYIESEETIEHSVENQSSSSSSSSSTIW 180
DB 80 PPSGASQITKSTKRSPPKPKKTKKYIESEETIEHSVENQSSSSSSSSSTIW 115
QY 181 KIKSSKNSANRELOKLLKLVONKKNRTPKPKPVPVVDAGSGLDNGDFKVTTPDTST 240
DB 116 -----VDONKKNRTPKPKPVPVVDAGSGLDNGDFKVTTPDTST 156
QY 241 TQHNKVSPTKITTAKPINRPSLPNSDTSKETSILVKNKETEYVETTTNNQSTDG 300
DB 157 TQHNKVSPTKITTAKPINRPSLPNSDTSKETSILVKNKETEYVETTTNNQSTDG 216
QY 301 KEKTSKAKETOSIKTSKOLAPTSKYLAKTTPKAEITTKGPALTTPKEPTPTPKEPAS 360
DB 217 KEKTSKAKETOSIKTSKOLAPTSKYLAKTTPKAEITTKGPALTTPKEPTPTPKEPAS 276
QY 361 TTPKEPTPTTKSAPTTPKEPAPTTTTSAPTTPKEPAPTTTKEBPATTTKEBPATTTKEP 420
DB 277 TTPKEPTPTTKSAPTTPKEPAPTTTTSAPTTPKEPAPTTTKEBPATTTKEBPATTTKEP 336
QY 421 APTTTSKAPTTPKEPAPTTPKKAPATTPKKBPATTTPKKBPATTTPKKBPATTTPK 480
DB 337 APTTTSKAPTTPKEPAPTTPKKAPATTPKKBPATTTPKKBPATTTPKKBPATTTPK 396
QY 481 EPAPTAPKKAPATTPKKBPATTTPKKBPATTTPKKBPATTTPKKBPATTTPKKBPAT 540
DB 397 EPAPTAPKKAPATTPKKBPATTTPKKBPATTTPKKBPATTTPKKBPATTTPKKBPAT 456
QY 541 TTKSAPTTPKKBPATTTPKKBPATTTPKKBPATTTPKKBPATTTPKKBPATTTPK 600
DB 457 TTKSAPTTPKKBPATTTPKKBPATTTPKKBPATTTPKKBPATTTPKKBPATTTPK 516
QY 601 APAPKBPATTTPKKBPATTTPKKBPATTTPKKBPATTTPKKBPATTTPKKBPAT 660
DB 517 APAPKBPATTTPKKBPATTTPKKBPATTTPKKBPATTTPKKBPATTTPKKBPAT 576
QY 661 PEEBPATTPKAAANTPEKBPATTTPKKBPATTTPKKBPATTTPKKBPATTTPKKBPAT 720
DB 577 PEEBPATTPKAAANTPEKBPATTTPKKBPATTTPKKBPATTTPKKBPATTTPKKBPAT 636
QY 721 APTTTPKAPKELAPTTTKEPTSTSDKBPATTTKGAATTTPKKBPATTTPKKBPAT 780
DB 637 APTTTPKAPKELAPTTTKEPTSTSDKBPATTTKGAATTTPKKBPATTTPKKBPAT 696
QY 781 TAPTTLKEBPATTTKBPAPKELAPTTTKEPTSTSDKBPATTTKGAATTTPKKBPAT 840
DB 697 TAPTTLKEBPATTTKBPAPKELAPTTTKEPTSTSDKBPATTTKGAATTTPKKBPAT 756
QY 841 KPAPTTTETPTPTTSEVSTPTTKEPTTIHKSPPESSTPELSAETPKALENSPREPGVT 900
DB 757 KPAPTTTETPTPTTSEVSTPTTKEPTTIHKSPPESSTPELSAETPKALENSPREPGVT 816
QY 901 TKTPTATKPEMTTAKKTERDRTTPETTTAPKMTKETATTTKEKTSKIRATTTQV 960
DB 817 TKTPTATKPEMTTAKKTERDRTTPETTTAPKMTKETATTTKEKTSKIRATTTQV 876
QY 961 TSTTTQDTTPPKITTLKTTTLAPKVTTTKITTTTTEIMNKEETAKPKDRATNSKATTPK 1020
DB 877 TSTTTQDTTPPKITTLKTTTLAPKVTTTKITTTTTEIMNKEETAKPKDRATNSKATTPK 936
QY 1021 POKPTAPKKPTSTKSKKTMKPRVAKKTTTPPKMTSTMBELNPTSIAEMLQTTTRPN 1080
DB 937 POKPTAPKKPTSTKSKKTMKPRVAKKTTTPPKMTSTMBELNPTSIAEMLQTTTRPN 996
QY 1081 QTPNSKLVENPKSEDEGAGETPHMLLRPHVMPVETDMOVLPRVNOGIIINPMLS 1140
DB 997 QTPNSKLVENPKSEDEGAGETPHMLLRPHVMPVETDMOVLPRVNOGIIINPMLS 1056
QY 1141 DETNINCKGKPVDTTLRNGTLVAFRGHYFMMLSPSPSPBARRITEVWGIPTDITVFT 1200
DB 1057 DETNINCKGKPVDTTLRNGTLVAFRGHYFMMLSPSPSPBARRITEVWGIPTDITVFT 1116

QY 1201 RCNCEGTFFPKDSQYWRFTNDIKDAGYPKPIKFGGGLTGOIUALSTAKYKNWPESEYV 1260
 DB 1117 RCNCEGTFFPKDSQYWRFTNDIKDAGYPKPIKFGGGLTGOIUALSTAKYKNWPESEYV 1176
 QY 1261 FFRGGSIOQYIYKOEBOVKCPGRPALNPVYGEOTVARRRFEPAIGPSQHTIRIQY 1320
 DB 1177 FFRGGSIOQYIYKOEBOVKCPGRPALNPVYGEOTVARRRFEPAIGPSQHTIRIQY 1236
 QY 1321 SPARLAVODKGVLANEYKVSILWRGLPNVTSAISLPNIKPGYDYAASSKOQYNNIDV 1380
 DB 1237 SPARLAVODKGVLANEYKVSILWRGLPNVTSAISLPNIKPGYDYAASSKOQYNNIDV 1296
 QY 1381 PSRTARAITTRSGOTLSKVMYNCP 1404
 DB 1297 PSRTARAITTRSGOTLSKVMYNCP 1320

RESULT 8

US-10-124-557-42
 / Sequence 42, Application US/10124557
 / Publication NO. US20020137894A1

GENERAL INFORMATION:

APPLICANT: Turner, Katherine
 Clark, Stephen C.
 Jacobs, Kenneth
 Hewick, Rodney M.
 Geener, Thomas G.
 TITLE OF INVENTION: Megakaryocyte Stimulating Factors
 NUMBER OF SEQUENCES: 143
 CORRESPONDENCE ADDRESS:
 ADDRESS: Genetics Institute, Inc.
 STREET: 87 Cambridgepark Drive
 CITY: Cambridge
 STATE: Massachusetts
 COUNTRY: U.S.A.
 ZIP: 02140

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent in Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/10/124,557
 FILING DATE: 16-Apr-2002

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/643,502
 FILING DATE: 18-JAN-1991
 APPLICATION NUMBER: US 07/546,114
 FILING DATE: 29-JUN-1990
 APPLICATION NUMBER: US 07/457,196
 FILING DATE: 29-DEC-1989
 APPLICATION NUMBER: US 07/390,901
 FILING DATE: 08-AUG-1989

ATTORNEY/AGENT INFORMATION:

NAME: Gaert, Luann
 REGISTRATION NUMBER: 31,822
 REFERENCE/DOCKET NUMBER: GI 5190
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (617)876-1170
 TELEFAX: (617)876-5851

INFORMATION FOR SEQ ID NO: 42:

SEQUENCE CHARACTERISTICS:
 LENGTH: 1311 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 SEQUENCE DESCRIPTION: SEQ ID NO: 42:
 US-10-124-557-42

Query Match 93.4%; Score 7007.5; DB 14; Length 1311;
 Best Local Similarity 93.4%; Pred. No. 0;

Matches 1311; Conservative 0; Mismatches 0; Indels 93; Gaps 1;
 QY 1 MAMKTIPIYLLILLISFVYIQOVSQDLSACAGCGGYSRDATCNDVNCQHYMECCPPD 60
 DB 1 MAMKTIPIYLLILLISFVYIQOVSQDLSACAGCGGYSRDATCNDVNCQHYMECCPPD 60
 QY 61 KAVCTAELSCGRCPESFERGRCDCDAQCKKYDKCCPYEAPSCAEVNHPTSPSSKAP 120
 DB 61 KAVCTAELSCGRCPESFERGRCDCDAQCKKYDKCCPYEAPSCAEVNHPTSPSSKAP 106
 QY 121 PPSGASQRIKSTTKKSPKPNKKTKVIESEITEHSEVSENOESSSSSSSSSIW 180
 DB 107 ----- 106
 QY 181 KIKSSKNSAANBELQKLVKONKKNRTKKKTPKPPVVDAGSGIDNDGFKYTPDTS 240
 DB 107 -----VKONKKNRTKKKTPKPPVVDAGSGIDNDGFKYTPDTS 147
 QY 241 TOHNKVSTPKITTAKPINRPSLPNSDTSKETSITVNEKETVETKETTNNKOTSDG 300
 DB 148 TOHNKVSTPKITTAKPINRPSLPNSDTSKETSITVNEKETVETKETTNNKOTSDG 207
 QY 301 KEKTSASKEQSIKETSANDLAPTSKVLAKEPTTKGPAITTPKEPTPTTPKEPAS 360
 DB 208 KEKTSASKEQSIKETSANDLAPTSKVLAKEPTTKGPAITTPKEPTPTTPKEPAS 267
 QY 361 TTPKEPTPTTISAPITTPKEPAITTPKSAITTPKEPAITTPKEPAITTPKEPA 420
 DB 268 TTPKEPTPTTISAPITTPKEPAITTPKSAITTPKEPAITTPKEPAITTPKEPA 327
 QY 421 APTTKSAITTPKEPAITTPKKAAPITTPKEPAITTPKEPTPTTPKEPAITTPKE 480
 DB 328 APTTKSAITTPKEPAITTPKKAAPITTPKEPAITTPKEPTPTTPKEPAITTPKE 387
 QY 481 EPAPTAAPKKAAPITTPKEPAITTPKEPAITTPKEPSPTTPKEPAITTPKSAITTP 540
 DB 388 EPAPTAAPKKAAPITTPKEPAITTPKEPAITTPKEPSPTTPKEPAITTPKSAITTP 447
 QY 541 TTKSAITTPKEPSPTTPKEPAITTPKEPAITTPKKAAPITTPKEPAITTPKEPA 600
 DB 448 TTKSAITTPKEPSPTTPKEPAITTPKEPAITTPKKAAPITTPKEPAITTPKEPA 507
 QY 601 APTAPKEPAITTPKETAAPITTPKKLPTTPKEKALPTTPKEPAITTPKEBAPITTP 660
 DB 508 APTAPKEPAITTPKETAAPITTPKKLPTTPKEKALPTTPKEPAITTPKEBAPITTP 567
 QY 661 PEPAPITTPKAAAPITTPKEPAITTPKEPAITTPKETAAPITTPKEPAITTPK 720
 DB 568 PEPAPITTPKAAAPITTPKEPAITTPKEPAITTPKETAAPITTPKEPAITTPK 627
 QY 721 APTTPKKAAPKEAPITTPKEPTSTSDKAPITTPKGAITTPKEPAITTPKEPA 780
 DB 628 APTTPKKAAPKEAPITTPKEPTSTSDKAPITTPKGAITTPKEPAITTPKEPA 687
 QY 781 TAPITLKEPAITTPKKAAPKEAPITTPKGAITTPKSTSDKAPITTPKETAAPITTP 840
 DB 688 TAPITLKEPAITTPKKAAPKEAPITTPKGAITTPKSTSDKAPITTPKETAAPITTP 747
 QY 841 KPAITTPETPTPTSEVSTPTTPKEPTTIHKSPTSSTPELSAAPTPKALENSPKERG 900
 DB 748 KPAITTPETPTPTSEVSTPTTPKEPTTIHKSPTSSTPELSAAPTPKALENSPKERG 807
 QY 901 TKTPAATPEMTTIAKDTTERDLRTTBETTTAAKMTKEITATTEKTTESKITATTTQV 960
 DB 808 TKTPAATPEMTTIAKDTTERDLRTTBETTTAAKMTKEITATTEKTTESKITATTTQV 867
 QY 961 TSTTQDTPPKITTLKTTTLAPKYVTTTKKITTEINMKPESTAKXORANNSATTPK 1020
 DB 868 TSTTQDTPPKITTLKTTTLAPKYVTTTKKITTEINMKPESTAKXORANNSATTPK 927
 QY 1021 POKPTKAPKKPTSTKPKTPMVRKPTTPPRKMTSTMBELNPSRIAEMALQTTTREN 1080
 DB 928 POKPTKAPKKPTSTKPKTPMVRKPTTPPRKMTSTMBELNPSRIAEMALQTTTREN 987

QY 1081 QTPNSKLVENVNPKSDAGAGETPHMLLRPHVMEVTPMDYLPRVNOGIIINPMLS 1140
DB 988 QTPNSKLVENVNPKSDAGAGETPHMLLRPHVMEVTPMDYLPRVNOGIIINPMLS 1047
QY 1141 DETNINCKGPDVGLTLLANGTLVAFRGHYFMWLSPPSPARRITTEVMGIPSPIDTVFT 1200
DB 1048 DETNINCKGPDVGLTLLANGTLVAFRGHYFMWLSPPSPARRITTEVMGIPSPIDTVFT 1107
QY 1201 RCNCEGKTFPRKDSQYMFNDIKDAGYPKPIFKFGGLTGQIVAASTAKYKMPBSVY 1260
DB 1108 RCNCEGKTFPRKDSQYMFNDIKDAGYPKPIFKFGGLTGQIVAASTAKYKMPBSVY 1167
QY 1261 FFKGGSIIQYIYKOEPOVKCPGRPALNYPYGEOTVRRRRFERAIGPSOTHTIRIQY 1320
DB 1168 FFKGGSIIQYIYKOEPOVKCPGRPALNYPYGEOTVRRRRFERAIGPSOTHTIRIQY 1227
QY 1321 SPARLAYODKVLHNEVKVSLIMRGLPNVVTSAISLPNIRKPDGYDYAFSKDQYNNIDV 1380
DB 1228 SPARLAYODKVLHNEVKVSLIMRGLPNVVTSAISLPNIRKPDGYDYAFSKDQYNNIDV 1287
QY 1381 PSRTARAITTSQGLSKVWYNCP 1404
DB 1288 PSRTARAITTSQGLSKVWYNCP 1311

RESULT 9
US-10-124-557-50
Sequence 50, Application US/10124557
Publication No. US20020137894A1
GENERAL INFORMATION:
APPLICANT: Turner, Katherine
Clark, Stephen C.
Jacobs, Kenneth
Hewick, Rodney M.
Gesner, Thomas G.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/124.557
FILING DATE: 16-Apr-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cseiff, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 876-1170
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 50:
SEQUENCE CHARACTERISTICS:
LENGTH: 1314 amino acids

TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 50:
US-10-124-557-50
Query Match 92.6%; Score 6971; DB 14; Length 1314;
Best Local Similarity 93.6%; Pred. No. 0; Mismatches 0; Indels 90; Gaps 1;
Matches 1314; Conservative 0; Mismatches 0; Indels 90; Gaps 1;
QY 1 MAMKTLPIYLLLLSVPIQQVSSQDLSGCGRSGEGYSDATGNCNDYNGHYMECCDF 60
DB 1 MAMKTLPIYLLLLSVPIQQVSSQDLSGCGRSGEGYSDATGNCNDYNGHYMECCDF 60
QY 61 KRVCIAELSCGRCFESFERREDCDAQCKYDKCCPDYSPCAEVNPNFSPSSKAP 120
DB 61 KRVCIAELSCGRCFESFERREDCDAQCKYDKCCPDYSPCAEVNPNFSPSSKAP 66
QY 121 PPSGASQTIKSTYKSPAPNKKKKVIESEIIEHSVENQESSSSSSSSSSSTTW 180
DB 67 -----EHSVENQESSSSSSSSSSSSSTTW 90
QY 181 KIKSSKNSAANRELQKLVKDNKKNRTKKKPTPKPVVDAGSGLDNGDFKVTPTDST 240
DB 91 KIKSSKNSAANRELQKLVKDNKKNRTKKKPTPKPVVDAGSGLDNGDFKVTPTDST 150
QY 241 TQHNKYSTPKITAKINRPSLPNSDTSKETSIVNKTETVETKETTNNKQTSIDG 300
DB 151 TQHNKYSTPKITAKINRPSLPNSDTSKETSIVNKTETVETKETTNNKQTSIDG 210
QY 301 KEKTTSAKETOSIKETSAKDLAPSKYLAKPRAETTTGAPALTTPKEPTTPKEPAS 360
DB 211 KEKTTSAKETOSIKETSAKDLAPSKYLAKPRAETTTGAPALTTPKEPTTPKEPAS 270
QY 361 TTPKEPTTPKSAPTTPKEBAPTTPKSAPTTPKEBAPTTPKSAPTTPKEBAPT 420
DB 271 TTPKEPTTPKSAPTTPKEBAPTTPKSAPTTPKEBAPTTPKSAPTTPKEBAPT 330
QY 421 APTTPKAPTPPKKAPATTPPKKAPATTPPKKAPATTPPKKAPATTPPKKAPAT 480
DB 331 APTTPKAPTPPKKAPATTPPKKAPATTPPKKAPATTPPKKAPATTPPKKAPAT 390
QY 481 EPAPTAKKKAPATTPPKKAPATTPPKKAPATTPPKKAPATTPPKKAPATTPPKKAP 540
DB 391 EPAPTAKKKAPATTPPKKAPATTPPKKAPATTPPKKAPATTPPKKAPATTPPKKAP 450
QY 541 TTKSAPTPPKESPPTTKKAPATTPPKKAPATTPPKKAPATTPPKKAPATTPPKKAP 600
DB 451 TTKSAPTPPKESPPTTKKAPATTPPKKAPATTPPKKAPATTPPKKAPATTPPKKAP 510
QY 601 APTAPKBPATTPKETAATTPPKKLTPTTPEKLAATTPPEKAPATTPPELAATTPPE 660
DB 511 APTAPKBPATTPKETAATTPPKKLTPTTPEKLAATTPPEKAPATTPPELAATTPPE 570
QY 661 PEEBAPTTPKAAANTPKKAPATTPPKKAPATTPPKKAPATTPPKKAPATTPPKKAP 720
DB 571 PEEBAPTTPKAAANTPKKAPATTPPKKAPATTPPKKAPATTPPKKAPATTPPKKAP 630
QY 721 APTTPKKAPKELAPTTTKEPTSTSDKAPATTPPKGAPATTPPKKAPATTPPKG 780
DB 631 APTTPKKAPKELAPTTTKEPTSTSDKAPATTPPKGAPATTPPKKAPATTPPKG 690
QY 781 TAPTTLEKAPATTPKAPKELAPTTTGSTSTSDKAPATTPPKETAATTPPKKAPATTP 840
DB 691 TAPTTLEKAPATTPKAPKELAPTTTGSTSTSDKAPATTPPKETAATTPPKKAPATTP 750
QY 841 KPAPTTETPTPTTSEVSTPTTKEPTTIKSPDSESTPELSABETPALENSPEKPVPT 900
DB 751 KPAPTTETPTPTTSEVSTPTTKEPTTIKSPDSESTPELSABETPALENSPEKPVPT 810
QY 901 TKTPPAATKPEMTTAKDKTTERDLRTTPETTTAAKPKATKETAATTEKTESKITATTT 960
DB 811 TKTPPAATKPEMTTAKDKTTERDLRTTPETTTAAKPKATKETAATTEKTESKITATTT 870

QY 961 TSTTODTTPKITTLLKTTTLLAKPVTYTKKITTTEIMNKPEETAKPKDRATNSKATTPK 1020
 DB 871 TSTTODTTPKITTLLKTTTLLAKPVTYTKKITTTEIMNKPEETAKPKDRATNSKATTPK 930
 QY 1021 POKPTAPKPKPTSTKPKPTMPRVKPKPTTPPKMTSTMBELNPTSRIAEAMLQTTTRN 1080
 DB 931 POKPTAPKPKPTSTKPKPTMPRVKPKPTTPPKMTSTMBELNPTSRIAEAMLQTTTRN 990
 QY 1081 QTPNSKLVENPKSEBAGAEGETPHMLRPHFMPEVTPDMOYLRLVRPNOGIIINPMIS 1140
 DB 991 QTPNSKLVENPKSEBAGAEGETPHMLRPHFMPEVTPDMOYLRLVRPNOGIIINPMIS 1050
 QY 1141 DETNINCKRPVDLTLTRNGTLVAFRGHYFMMLSPSPSPARITREWGIPSPIDTVP 1200
 DB 1051 DETNINCKRPVDLTLTRNGTLVAFRGHYFMMLSPSPSPARITREWGIPSPIDTVP 1110
 QY 1201 RCNCEGKTFPPKDSQYWRFTNDIKDAGYPRPKFGGLTGQIVAAALSTAKYKMPESY 1260
 DB 1111 RCNCEGKTFPPKDSQYWRFTNDIKDAGYPRPKFGGLTGQIVAAALSTAKYKMPESY 1170
 QY 1261 PFRGGSIOQYIYKQSPVQKCPGRPALNYPVYGEHTQVRRRFRERAIQPSQTHTRIOY 1320
 DB 1171 PFRGGSIOQYIYKQSPVQKCPGRPALNYPVYGEHTQVRRRFRERAIQPSQTHTRIOY 1230
 QY 1321 SPARLAYODKGVLANEYKVSILWRGLPNVYTSALSPNIRKPDGYDYAFSKOQYNNIDV 1380
 DB 1231 SPARLAYODKGVLANEYKVSILWRGLPNVYTSALSPNIRKPDGYDYAFSKOQYNNIDV 1290
 QY 1381 PSRTARAITTRSGQTLISKWYNCP 1404
 DB 1291 PSRTARAITTRSGQTLISKWYNCP 1314

RESULT 10

US-10-124-557-142
 Sequence 142, Application US/10124557
 Publication No. US20020137894A1

GENERAL INFORMATION:

APPLICANT: Turner, Katherine
 Clark, Stephen C.
 Jacobs, Kenneth
 Hewick, Rodney M.
 Geener, Thomas G.
 TITLE OR INVENTION: Megakaryocyte Stimulating Factors
 NUMBER OF SEQUENCES: 143
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Genetice Institute, Inc.
 STREET: 87 Cambridgepark Drive
 CITY: Cambridge
 STATE: Massachusetts
 COUNTRY: U.S.A.
 ZIP: 02140

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10124,557
 FILING DATE: 16-Apr-2002
 CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/643,502
 FILING DATE: 18-JAN-1991
 APPLICATION NUMBER: US 07/546,114
 FILING DATE: 29-JUN-1990
 APPLICATION NUMBER: US 07/457,196
 FILING DATE: 29-DEC-1989
 APPLICATION NUMBER: US 07/390,901
 FILING DATE: 08-AUG-1989
 ATTORNEY/AGENT INFORMATION:
 NAME: Ceebr, Luann

REGISTRATION NUMBER: 31,822
 REFERENCE/DOCKET NUMBER: GI 5190
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (617) 876-1170
 TELEFAX: (617) 876-5851
 INFORMATION FOR SEQ ID NO: 142:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1313 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 SEQUENCE DESCRIPTION: SEQ ID NO: 142:
 US-10-124-557-142

Query Match 92.4%; Score 6951.5; DB 14; Length 1313;
 Best Local Similarity 93.5%; Pred. No. 0;
 Matches 1313; Conservative 0; Mismatches 0; Indels 91; Gaps 2;

1 MAMKTLPIYLILLLVFVLIQVVSQDLSSACAGCEGYSRDATCNCNDVQCQHYMECCPDF 60
 DB 1 MAMKTLPIYLILLLVFVLIQVVSQ----- 25
 QY 61 KKVCTAELSCKRCESFERGECDDAQCKKYDKCCPDYSFCAEVHNPTSPPSSKAP 120
 DB 26 -----ELSCKRCESFERGECDDAQCKKYDKCCPDYSFCA----- 64
 QY 121 PPSGASQITTKTSKSPRPKTKKVIIESEITEHSVSNQSSSSSSSSSSSSSTTW 180
 DB 65 -----EEHSVSNQSSSSSSSSSSSSSSSTTW 89
 QY 181 KIKSSKNSAANRELQKLVKNDKKNRTKKKTPKPVVDAGSGLDNDPFVYTPDPST 240
 DB 90 KIKSSKNSAANRELQKLVKNDKKNRTKKKTPKPVVDAGSGLDNDPFVYTPDPST 149
 QY 241 TQHNKVSSTPKITTKAPINRPSPNSDTSKETSIVNKEITVEKETTNNKQSTIDG 300
 DB 150 TQHNKVSSTPKITTKAPINRPSPNSDTSKETSIVNKEITVEKETTNNKQSTIDG 209
 QY 301 KEKTTSAEQTQSIETKSADLAPTSKVLAKPPKAEITTTKGALTTTPKEPTTTKEPAS 360
 DB 210 KEKTTSAEQTQSIETKSADLAPTSKVLAKPPKAEITTTKGALTTTPKEPTTTKEPAS 269
 QY 361 TTPKEPTTTIKSAPTTPKEBAPTTPKSAPTTPKBPAPTTPKBPAPTTPKBPAPTTPK 420
 DB 270 TTPKEPTTTIKSAPTTPKEBAPTTPKSAPTTPKBPAPTTPKBPAPTTPKBPAPTTPK 329
 QY 421 APTTKSAPTTPKEBAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPK 480
 DB 330 APTTKSAPTTPKEBAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPK 389
 QY 481 BPAPTAAPKPAPTTPKEBAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPT 540
 DB 390 BPAPTAAPKPAPTTPKEBAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPT 449
 QY 541 TTKSAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPT 600
 DB 450 TTKSAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPT 509
 QY 601 APTAPKEBAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPK 660
 DB 510 APTAPKEBAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPK 569
 QY 661 PEEBAPTTPKAAAPNTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPT 720
 DB 570 PEEBAPTTPKAAAPNTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPT 629
 QY 721 APTTPKBPAPKELAPTTTKEPTSTSDKAPPTTGGTAPTPPKBPAPTTPKBPAPTTPKBPAPT 780
 DB 630 APTTPKBPAPKELAPTTTKEPTSTSDKAPPTTGGTAPTPPKBPAPTTPKBPAPTTPKBPAPT 689
 QY 781 TAPTLKEBAPTTPKBPAPKELAPTTTGGTAPTTSDKAPPTTPKBPAPTTPKBPAPTTPKBPAPT 840
 DB 690 TAPTLKEBAPTTPKBPAPKELAPTTTGGTAPTTSDKAPPTTPKBPAPTTPKBPAPTTPKBPAPT 749

QY 841 KPAPTPPTPTSEVSTPTTKKPTTIHKSPESTBELSAEPKALENSPEKPGPT 900
DB 750 KPAPTPPTPTSEVSTPTTKKPTTIHKSPESTBELSAEPKALENSPEKPGPT 809
QY 901 TKTPAATPEMTTAKOKTERDRLATTPETTTAAAPKMTKETATTTTEKTESKITATTTQV 960
DB 810 TKTPAATPEMTTAKOKTERDRLATTPETTTAAAPKMTKETATTTTEKTESKITATTTQV 869
QY 961 TSTTTODTTPPKITTLKTTTLAPKVTYTKITITTEINAKKEBETAKPKORATNSKATTPK 1020
DB 870 TSTTTODTTPPKITTLKTTTLAPKVTYTKITITTEINAKKEBETAKPKORATNSKATTPK 929
QY 1021 POKTPKAPKPTSTKPKTPMVRKPKTPPTPKMTSTMPBELNPTSLAEMLOTTTREN 1080
DB 930 POKTPKAPKPTSTKPKTPMVRKPKTPPTPKMTSTMPBELNPTSLAEMLOTTTREN 989
QY 1081 QTPNSKIVEVNPKESEDAGAGETPHMLLRPHVPEVTPMDYLLPRVNOGIIINPMLS 1140
DB 990 QTPNSKIVEVNPKESEDAGAGETPHMLLRPHVPEVTPMDYLLPRVNOGIIINPMLS 1049
QY 1141 DETNINCKGKPYDGLTTLNGLTVAFRGHYFWMLSFSPSPSPARRITTEVWGIIPSPIDTFT 1200
DB 1050 DETNINCKGKPYDGLTTLNGLTVAFRGHYFWMLSFSPSPSPARRITTEVWGIIPSPIDTFT 1109
QY 1201 RCNCEGKTFEPKDSQYMFRTNDIKDAGYKPIPKGFGGLTGQIYVAALSTAKYKMMPESSV 1260
DB 1110 RCNCEGKTFEPKDSQYMFRTNDIKDAGYKPIPKGFGGLTGQIYVAALSTAKYKMMPESSV 1169
QY 1261 FFKGSGSIQOYIYKQEPYQKCPGRPALNVPYGEVMTQVRRRRRPERAIGPSQTHIRIQY 1320
DB 1170 FFKGSGSIQOYIYKQEPYQKCPGRPALNVPYGEVMTQVRRRRRPERAIGPSQTHIRIQY 1229
QY 1321 SPAPLAYODKVLNEVAVSILMRGLPNVVTSAISLPIRKPDGYDYAFSKOQYNTIDV 1380
DB 1230 SPAPLAYODKVLNEVAVSILMRGLPNVVTSAISLPIRKPDGYDYAFSKOQYNTIDV 1289
QY 1381 PSRTARATITRSQGTLSKVMTNCP 1404
DB 1290 PSRTARATITRSQGTLSKVMTNCP 1313

RESULT 11
US-10-124-557-44
; Sequence 44, Application US/10124557
; Publication No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Clark, Stephen C.
; Jacobs, Kenneth
; Hewick, Rodney M.
; Gesner, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/124,557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991

APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Geert, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 876-1170
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 44:
SEQUENCE CHARACTERISTICS:
LENGTH: 1270 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 44:
US-10-124-557-44
Query Match 89.4%; Score 6727; DB 14; Length 1270;
Best Local Similarity: 90.5%; Pred. No. 0;
Matches 1270; Conservative 0; Mismatches 0; Indels 134; Gaps 2;
QY 1 MAMKTLPIYLLLSVIVIOVSSODLSSCAGRGEGSRDATCNQCHVMECCPDF 60
DB 1 MAMKTLPIYLLLSVIVIOVSSQ----- 25
QY 61 KRVCIAELSCKGRCFSEFERGECDCDQCKYDKCCPDYESFCAEVNPTSPSSKAP 120
DB 26 -----ELSCKGRCFSEFERGECDCDQCKYDKCCPDYESFCAE----- 65
QY 121 PPGASQITKSTTRSPKPKKKYKVIIESEITBEHSVENQSSSSSSSSSTIW 180
DB 66 ----- 65
QY 181 KIKSKNSAANRELQKLLKYNKKNKTKKKKTPKPVVNDAGSLNDGPKVTPPTST 240
DB 66 -----VDNKNKNTKKKPKPKPVVNDAGSLNDGPKVTPPTST 106
QY 241 TQHNKSTSPKLTAKINPRLPSLPNSDTSKETSITVNNKETTVEKETTNNKQSTDG 300
DB 107 TQHNKSTSPKLTAKINPRLPSLPNSDTSKETSITVNNKETTVEKETTNNKQSTDG 166
QY 301 KEKTTSAKETQSIEKTSAKDLAPTSLYAKPTPAETTTKGPALTTPKEPTTPKEPAS 360
DB 167 KEKTTSAKETQSIEKTSAKDLAPTSLYAKPTPAETTTKGPALTTPKEPTTPKEPAS 226
QY 361 TTPKEPTTPKSAPTTPKEPAPTTTSAPPTTPKEPAPTTTPKEPAPTTTPKEPA 420
DB 227 TTPKEPTTPKSAPTTPKEPAPTTTSAPPTTPKEPAPTTTPKEPAPTTTPKEPA 286
QY 421 APTTTSAPPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTT 480
DB 287 APTTTSAPPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTT 346
QY 481 EPAPTAPEKPAPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPA 540
DB 347 EPAPTAPEKPAPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPA 406
QY 541 TTKSAPPTTPKEPSPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTP 600
DB 407 TTKSAPPTTPKEPSPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTP 466
QY 601 APTAPKEPAPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPT 660
DB 467 APTAPKEPAPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPT 526
QY 661 PEEBAPPTPKAAADNTKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPA 720
DB 527 PEEBAPPTPKAAADNTKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPA 586

```

QY 721 APTPKKPAKELAPTTKEPTSTSDKPAPTPKGTAPTPKPAPTPKKPAPTPKG 780
| | | | |
DB 587 APTPKKPAKELAPTTKEPTSTSDKPAPTPKGTAPTPKKPAPTPKKPAPTPKG 646
QY 781 TAPTLKEBAPTPPKKPAKELAPTTKGPTSTSDKPAPTPKETAPTPKKPAPTPK 840
| | | | |
DB 647 TAPTLKEBAPTPPKKPAKELAPTTKGPTSTSDKPAPTPKETAPTPKKPAPTPK 706
QY 841 KAPPTPEPTPTTSVSPTTKEPTTHKSDESTPELSAPTPKALENSKEGVP 900
| | | | |
DB 707 KAPPTPEPTPTTSVSPTTKEPTTHKSDESTPELSAPTPKALENSKEGVP 766
QY 901 TTPATKPEMTTAKDKTERDLATTPETTAAPKMTKETATTTKTESKITATTTVOY 960
| | | | |
DB 767 TTPATKPEMTTAKDKTERDLATTPETTAAPKMTKETATTTKTESKITATTTVOY 826
QY 961 TSTTODTPPKITTLKTTTLADKVTYTKKTTTTEIMNKPEETAKPKDRAATSKATTPK 1020
| | | | |
DB 827 TSTTODTPPKITTLKTTTLADKVTYTKKTTTTEIMNKPEETAKPKDRAATSKATTPK 886
QY 1021 POKPTAPKPKPTSTKPKTPMRAKPKTTPPKKMTSTMBELNPTSRILAMLOTTTRN 1080
| | | | |
DB 887 POKPTAPKPKPTSTKPKTPMRAKPKTTPPKKMTSTMBELNPTSRILAMLOTTTRN 946
QY 1081 QTPNSKLVEVNPKESEDAGAGEETPHMLLRPHVFMPEVTPDMOYLPRVFNQGIINPMLS 1140
| | | | |
DB 947 QTPNSKLVEVNPKESEDAGAGEETPHMLLRPHVFMPEVTPDMOYLPRVFNQGIINPMLS 1006
QY 1141 DETNINCKRPVDGLTTLRNGTTLVAFRGHYFMMLSPSPSPARITTEWGISPIDTVP 1200
| | | | |
DB 1007 DETNINCKRPVDGLTTLRNGTTLVAFRGHYFMMLSPSPSPARITTEWGISPIDTVP 1066
QY 1201 RNCBCKTFFPKDSQVWRFTNDIKDAGYKPKFKGSGGLTGOVALSTAKTKMVESTY 1260
| | | | |
DB 1067 RNCBCKTFFPKDSQVWRFTNDIKDAGYKPKFKGSGGLTGOVALSTAKTKMVESTY 1126
QY 1261 PFKRGSIOQYIYKQBPVOKCPGRPALNYPVGEVTOVRRRRFEALIGPSQHTIRIY 1320
| | | | |
DB 1127 PFKRGSIOQYIYKQBPVOKCPGRPALNYPVGEVTOVRRRRFEALIGPSQHTIRIY 1186
QY 1321 SPARLAQDKGVLANEVKVSILMRGLPNVYTSALISLPIRIKPDGYDYVAFSKQOYINIV 1380
| | | | |
DB 1187 SPARLAQDKGVLANEVKVSILMRGLPNVYTSALISLPIRIKPDGYDYVAFSKQOYINIV 1246
QY 1381 PSRTAAITTRSGQITLSKVMYVNC 1404
| | | | |
DB 1247 PSRTAAITTRSGQITLSKVMYVNC 1270

```

RESULT 12

```

US-10-124-557-104
; Sequence 104, Application US/10124557
; Publication No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
;           Jacob, Stephen C.
;           Jacob, Kenneth
;           Hewick, Rodney M.
;           Geener, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

```

```

; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/124,557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Geert, Luam
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 104:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1140 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 104:
US-10-124-557-104

Query Match      80.9%; Score 6085; DB 14; Length 1140;
Best Local Similarity 100.0%; Pred. No. 5,86-302;
Matches 1140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1  MAMKTLPIYLILLISLVFIYIQVSSQDLSACAGRCGEYSRDATCNDYNCQHYMECCPDF 60
1  MAMKTLPIYLILLISLVFIYIQVSSQDLSACAGRCGEYSRDATCNDYNCQHYMECCPDF 60
61  KAVCTAELSCRCRCESEFERGECDDAQCKKYDCCPDYSEFCALVHNPTSPSSKAP 120
61  KAVCTAELSCRCRCESEFERGECDDAQCKKYDCCPDYSEFCALVHNPTSPSSKAP 120
121  PPSGASQTIKSTTKSPKPPNKKTKYIESBEITEBHSVSENOSSSSSSSSSSSTIM 180
121  PPSGASQTIKSTTKSPKPPNKKTKYIESBEITEBHSVSENOSSSSSSSSSSSTIM 180
181  KIKSSKNSAANRELQKLLKVDNKKKRTKKKTPKRPVVDAGSGLDNGDFKVTTPDTS 240
181  KIKSSKNSAANRELQKLLKVDNKKKRTKKKTPKRPVVDAGSGLDNGDFKVTTPDTS 240
241  TQHNKVSTSPKITTAKPINRPSLPPNSDTSKETSILTVNKEITVETKETTTNNKQTSIDG 300
241  TQHNKVSTSPKITTAKPINRPSLPPNSDTSKETSILTVNKEITVETKETTTNNKQTSIDG 300
301  KEKTSABKQSIKETSADLAPTSKYLAKPPKATTTTGGPALTTPKBPPTTPKEPAS 360
301  KEKTSABKQSIKETSADLAPTSKYLAKPPKATTTTGGPALTTPKBPPTTPKEPAS 360
361  TTPKEPTPTTKSAPTPPKKPAPTTKKSAPTTPPKKPAPTTKKPAPTTKKPAPTTKK 420
361  TTPKEPTPTTKSAPTPPKKPAPTTKKSAPTTPPKKPAPTTKKPAPTTKKPAPTTKK 420
421  APTTKSAPTPKKPAPTTKKPAPTTKKSAPTTPPKKPAPTTKKSAPTTKKPAPTTKK 480
421  APTTKSAPTPKKPAPTTKKPAPTTKKSAPTTPPKKPAPTTKKSAPTTKKPAPTTKK 480
481  BPAPPAKPKAPATTPKPAPTTKKPAPTTKKSAPTTPPKKPAPTTKKSAPTTKKPA 540
481  BPAPPAKPKAPATTPKPAPTTKKPAPTTKKSAPTTPPKKPAPTTKKSAPTTKKPA 540
541  TTKSAPTPKKBSPTTKKPAPTTKKPAPTTKKPAPTTKKPAPTTKKPAPTTKKPA 600
541  TTKSAPTPKKBSPTTKKPAPTTKKPAPTTKKPAPTTKKPAPTTKKPAPTTKKPA 600

```

QY 601 APTAPKBPAPPTPKETAPPTPKLPTTPKLAAPTPKPAAPTPPEELAPTPPEPTT 660
DB 601 APTAPKBPAPPTPKETAPPTPKLPTTPKLAAPTPKPAAPTPPEELAPTPPEPTT 660
QY 661 PEEBAPPTPKAAAPTPKPAAPTPPKBAPPTPKBAPPTPKETAPTPKGAPTTLKEP 720
DB 661 PEEBAPPTPKAAAPTPKPAAPTPPKBAPPTPKBAPPTPKETAPTPKGAPTTLKEP 720
QY 721 APTTPKAPKELAPPTTKBPTSTSDKAPPTPKGAPTPPKGAPTPPKBAPPTPKG 780
DB 721 APTTPKAPKELAPPTTKBPTSTSDKAPPTPKGAPTPPKGAPTPPKBAPPTPKG 780
QY 781 TAPPTLKBPAPPTPKKAPKELAPPTTKGPTSTSDKAPPTPKETAPTTPKPAAPTPK 840
DB 781 TAPPTLKBPAPPTPKKAPKELAPPTTKGPTSTSDKAPPTPKETAPTTPKPAAPTPK 840
QY 841 KPAPTPPTPTTSEVSTPTTKBPTTHKS PDESTPELSAAPTPKALENS PKBGPVT 900
DB 841 KPAPTPPTPTTSEVSTPTTKBPTTHKS PDESTPELSAAPTPKALENS PKBGPVT 900
QY 901 TKTPATPEMTTAKDKTERDLPTTETTTAAPTMTKETATTTTEKTESKITATTTTQV 960
DB 901 TKTPATPEMTTAKDKTERDLPTTETTTAAPTMTKETATTTTEKTESKITATTTTQV 960
QY 961 TSTTTODTPPKITLTKTTLLAPKVTTKITTTTEIMNKDEPTAKPKDAPATNSKATTPK 1020
DB 961 TSTTTODTPPKITLTKTTLLAPKVTTKITTTTEIMNKDEPTAKPKDAPATNSKATTPK 1020
QY 1021 POKPTPAKPKETSTKPKTPMVRKPTTPTPKKATSTMPBLNPTSTIAEAMLOTTTTPN 1080
DB 1021 POKPTPAKPKETSTKPKTPMVRKPTTPTPKKATSTMPBLNPTSTIAEAMLOTTTTPN 1080
QY 1081 QTPNSKLVENVPKSGEDAGAGETPHMLLRHVPEPTTDMDYLPKPNOCIIINPMLS 1140
DB 1081 QTPNSKLVENVPKSGEDAGAGETPHMLLRHVPEPTTDMDYLPKPNOCIIINPMLS 1140

RESULT 13

US-10-124-557-58
Sequence 58, Application US/10124557
Publication No. US20020137894A1
GENERAL INFORMATION:
APPLICANT: Turner, Katherine
Clark, Stephen C.
Jacobs, Kenneth
Hewick, Rodney M.
Gesner, Thomas G.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridge Park Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/124,557
FILING DATE: 16-Apr-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989

APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cseri, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 876-1170
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 58:
SEQUENCE CHARACTERISTICS:
LENGTH: 1049 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 58:
US-10-124-557-58
Query Match 73.2%; Score 5510.5; DB 14; Length 1049;
Best Local Similarity 92.0%; Pred. No. 9.5e-273;
Matches 1049; Conservative 0; Mismatches 0; Indels 91; Gaps 2;
QY 1 MAMKTLPIYLLLSVFVIOQVSSQDLSACAGCGEGYSRDATGNCQYHMECCPDF 60
DB 1 MAMKTLPIYLLLSVFVIOQVSSQ----- 25
QY 61 KRVTALSCGGRFESFERGECDDAOCCKYDKCCPDYESFCALYHNPSTPSSSKAP 120
DB 26 -----ELSCGGRFESFERGECDDAOCCKYDKCCPDYESFCA----- 64
QY 121 PPSGASQTIKSTGRSPKPKKTKKIVISEEITHEHSVENOESSSSSSSSSSSIW 180
DB 65 -----EHSVENOESSSSSSSSSSSIW 89
QY 181 KIKSSKNSANRELQKLVKDNKKNRTKKKPKPVVDAGSLDNGDFKVTTPDTST 240
DB 90 KIKSSKNSANRELQKLVKDNKKNRTKKKPKPVVDAGSLDNGDFKVTTPDTST 149
QY 241 TQHNKVSPTKITTAKPINRPSLPNSDTSKETSILVNKETTETKETTNNKQTSIDG 300
DB 150 TQHNKVSPTKITTAKPINRPSLPNSDTSKETSILVNKETTETKETTNNKQTSIDG 209
QY 301 KEKTTSAKETQSIEKTSAKDLAPSKVLAKPTPAETTTKGPALTTPKEPTTPKBPAS 360
DB 210 KEKTTSAKETQSIEKTSAKDLAPSKVLAKPTPAETTTKGPALTTPKEPTTPKBPAS 269
QY 361 TTPKEPTPTTIKSAPTTPKBPAPTTTKSAPTTPKBPAPTTTKBPAPTTTKBP 420
DB 270 TTPKEPTPTTIKSAPTTPKBPAPTTTKSAPTTPKBPAPTTTKBPAPTTTKBP 329
QY 421 APTTKSAPTPKBPAPTTPKKAPPTPKBAPPTPKBAPPTPKBAPTTTKBPAPTTPK 480
DB 330 APTTKSAPTPKBPAPTTPKKAPPTPKBAPPTPKBAPPTPKBAPTTTKBPAPTTPK 389
QY 481 EPAPTAKKBPAPTPKBPAPTTPKBAPTTTKBSPPTPKBAPTTTKSAATTTKEBAPT 540
DB 390 EPAPTAKKBPAPTPKBPAPTTPKBAPTTTKBSPPTPKBAPTTTKSAATTTKEBAPT 449
QY 541 TTKSAPTPKBPSPPTTKBPAPTTPKBAPPTPKKAPPTPKBAPTTTKBPAPTTTKBP 600
DB 450 TTKSAPTPKBPSPPTTKBPAPTTPKBAPPTPKKAPPTPKBAPTTTKBPAPTTTKBP 509
QY 601 APTAPKBPAPTPKETAPTTPKBLATTPTEKAPPTPEELAPTPPEEPTT 569
DB 510 APTAPKBPAPTPKETAPTTPKBLATTPTEKAPPTPEELAPTPPEEPTT 569
QY 661 PEEBAPPTPKAAAPTPKPAAPTPPKBAPPTPKBAPPTPKETAPTPKGAPTTLKEP 720
DB 570 PEEBAPPTPKAAAPTPKPAAPTPPKBAPPTPKBAPPTPKETAPTPKGAPTTLKEP 629
QY 721 APTTPKAPKELAPPTTKBPTSTSDKAPPTPKGAPTPPKGAPTPPKBAPPTPKG 780
DB 630 APTTPKAPKELAPPTTKBPTSTSDKAPPTPKGAPTPPKGAPTPPKBAPPTPKG 689

QY 781 TAPTLKEBAPPTPKKAPKELAPTTKGTSTSDKAPPTPKETAPPTPKKAPPTPK 840
 DB 690 TAPTLKEBAPPTPKKAPKELAPTTKGTSTSDKAPPTPKETAPPTPKKAPPTPK 749
 QY 841 KCAPPTPEPTPTTSVSTPTTKEPTTIHKSPDESTPELSAPPTKALENSKPEGVPT 900
 DB 750 KCAPPTPEPTPTTSVSTPTTKEPTTIHKSPDESTPELSAPPTKALENSKPEGVPT 809
 QY 901 TTPPATKREMTTAAOKTERDLRTTPPTTAAPOKTEATTTTKESTKITATTTOY 960
 DB 810 TTPPATKREMTTAAOKTERDLRTTPPTTAAPOKTEATTTTKESTKITATTTOY 869
 QY 961 TSTTODTTPPKITLTKITTLAPKVTTKKTIITTEIIMKPEETAPKORATNSKATPK 1020
 DB 870 TSTTODTTPPKITLTKITTLAPKVTTKKTIITTEIIMKPEETAPKORATNSKATPK 929
 QY 1021 PKKPTAPKPKPTSTKPKTMPRVKPKPTTPPKMTSTWELNPTSRIAEMLQTTTRN 1080
 DB 930 PKKPTAPKPKPTSTKPKTMPRVKPKPTTPPKMTSTWELNPTSRIAEMLQTTTRN 989
 QY 1081 QTPNSKLVENPKSEBAGABEGTPTMLLRPHVFMEEVTPDMDYLRVNOGIIINPMIS 1140
 DB 990 QTPNSKLVENPKSEBAGABEGTPTMLLRPHVFMEEVTPDMDYLRVNOGIIINPMIS 1049

RESULT 14

US-10-124-557-84
 / Sequence 84, Application US/10124557
 / Publication No. US20020137894A1
 / GENERAL INFORMATION:

APPLICANT: Turner, Katherine
 Clark, Stephen C.
 Jacobs, Kenneth
 Hewick, Rodney M.
 Geener, Thomas G.

TITLE OF INVENTION: Megakaryocyte Stimulating Factors
 NUMBER OF SEQUENCES: 143

CORRESPONDENCE ADDRESS:

ADDRESSEE: Genetics Institute, Inc.
 STREET: 87 Cambridgepark Drive
 CITY: Cambridge
 STATE: Massachusetts
 COUNTRY: U.S.A.

ZIP: 02140

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/124,557

FILING DATE: 16-Apr-2002

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/643,502

FILING DATE: 18-JAN-1991

APPLICATION NUMBER: US 07/546,114

FILING DATE: 29-JUN-1990

APPLICATION NUMBER: US 07/457,196

FILING DATE: 29-DEC-1989

APPLICATION NUMBER: US 07/390,901

ATTORNEY/AGENT INFORMATION:

NAME: Geert, Luann

REGISTRATION NUMBER: 31,822

REFERENCE/DOCKET NUMBER: GI 5190

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617)876-1170

TELEFAX: (617)876-5851

INFORMATION FOR SEQ ID NO: 84:

SEQUENCE CHARACTERISTICS:

LENGTH: 1022 amino acids

TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 SEQUENCE DESCRIPTION: SEQ ID NO: 84:
 US-10-124-557-84

Query Match 72.1%; Score 5425.5; DB 14; Length 1022;
 Best Local Similarity 91.5%; Pred. No. 26-268;
 Matches 1020; Conservative 0; Mismatches 2; Indels 93; Gaps 2;

QY 26 DISSCAGRGEGSYSDATCNCDCYNCOHYMECCPDFKRVCTAELSCGRGCFSEBGRGEC 85
 DB 1 DISSCAGRGEGSYSDATCNCDCYNCOHYMECCPDFKRVCTAELSCGRGCFSEBGRGEC 58
 QY 86 CPAQCKKIDKCCPDYDSCAEVHNPSTSPSSKAPPGASQITSTYGRSKFPNKKKT 145
 DB 59 CPAQCKKIDKCCPDYDSCAEVHNPSTSPSSKAPPGASQITSTYGRSKFPNKKKT 80
 QY 146 KVISEETHEHVSSENOSSSSSSSSSSSTIKIKSKNSAANRELQKLYKDNKK 205
 DB 81 -----AVKDNKK 87
 QY 206 NRTKKKPKPPEVVDAGSGLDNGDFKYTTPDSTTQHNKYSTSPKITAKBINRPSLP 265
 DB 88 NRTKKKPKPPEVVDAGSGLDNGDFKYTTPDSTTQHNKYSTSPKITAKBINRPSLP 147
 QY 266 PMSDTSKETSLSLVNKEITVETKETTNNKQTSIDGKEKTSKAKTOSIEKTSKADLAPTS 325
 DB 148 PMSDTSKETSLSLVNKEITVETKETTNNKQTSIDGKEKTSKAKTOSIEKTSKADLAPTS 207
 QY 326 KVLAKPTKAEETTTGPAITTPKEPTPTTPKBPASTPKEPTTIKASAPTTPKKBPAPT 385
 DB 208 KVLAKPTKAEETTTGPAITTPKEPTPTTPKBPASTPKEPTTIKASAPTTPKKBPAPT 267
 QY 386 TKSAPTTKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPK 445
 DB 268 TKSAPTTKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPK 327
 QY 446 TTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPT 505
 DB 328 TTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPT 387
 QY 506 PAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPT 565
 DB 388 PAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPT 447
 QY 566 KKPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPK 625
 DB 448 KKPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPK 507
 QY 626 PTPPEKLAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPT 685
 DB 508 PTPPEKLAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPT 567
 QY 686 KKPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPK 745
 DB 568 KKPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPK 627
 QY 746 SDKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPT 805
 DB 628 SDKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPT 687
 QY 806 TTKGPTSTTSOKPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTP 865
 DB 688 TTKGPTSTTSOKPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTP 747
 QY 866 PTTIHKSPDESTPELSAPPTKALENSKPEGVPTTKPATKPEMTTAAOKTERDLR 925
 DB 748 PTTIHKSPDESTPELSAPPTKALENSKPEGVPTTKPATKPEMTTAAOKTERDLR 807
 QY 926 TTPETTTAAPKOTKETATTTKTESKITATTTOYTSSTTODTTPFKITLTKITTLAPKV 985
 DB 808 TTPETTTAAPKOTKETATTTKTESKITATTTOYTSSTTODTTPFKITLTKITTLAPKV 867

Qy	1141	DETNICGKPVDDGLTTLRNGTLVAFRGHYFMMLSPSPSPARITEVWGIPSPIDTVFT	1200
Db	1007	-----	1006
Qy	1201	RCNCEGKTFFPKDSQYWRFTNDIKDAGYPKPIFKGFGGLTGQIWAALSTAKYNWPESVY	1260
Db	1007	-----	1006
Qy	1261	FFRGGSIQOYIYKQEPVQKCPGRPALNYPVYGEMTOVRRRRFERAIGBSQTHIRIOY	1320
Db	1007	-----	1006
Qy	1321	SPARLAYQDKGVLANEVKVSILWRGLPNVVTSAISLPNIRKPDGYDYAFSKDQYNYNDY	1380
Db	1007	-----	1014
Qy	1381	PSRTARAITTRSGQTLSTKYWYNCP	1404
Db	1015	PSRTARAITTRSGQTLSTKYWYNCP	1036

Search completed: December 8, 2003, 09:43:03
 Job time : 51 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 8, 2003, 09:34:07 ; Search time 22 Seconds
(without alignments)
2700.202 Million cell updates/sec

Title: US-09-556-246-1

Perfect score: 7526

Sequence: 1 MAMKTLPIYLLLLSVFVIQ.....ARAITRSGQTLSKWVNCNP 1404

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

1: Issued Patents AA:*
2: /cgn2_6/prodata/1/1aa/58_COMB.pep:*
3: /cgn2_6/prodata/1/1aa/58_COMB.pep:*
4: /cgn2_6/prodata/1/1aa/68_COMB.pep:*
5: /cgn2_6/prodata/1/1aa/PCTUS_COMB.pep:*
6: /cgn2_6/prodata/1/1aa/Backfill681.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7526	100.0	1404	4	US-07-757-022B-2
2	7526	100.0	1404	4	US-07-757-022B-62
3	7291.5	96.9	1361	4	US-07-757-022B-40
4	7245.5	96.3	1363	4	US-07-757-022B-52
5	7232	96.1	1354	4	US-07-757-022B-48
6	7011	93.2	1320	4	US-07-757-022B-46
7	7011	93.2	1310	4	US-07-757-022B-42
8	7007.5	93.1	1311	4	US-07-757-022B-40
9	6971	92.6	1314	4	US-07-757-022B-50
10	6951.5	92.4	1313	4	US-07-757-022B-142
11	6727	89.4	1270	4	US-07-757-022B-44
12	6085	80.9	1140	4	US-07-757-022B-104
13	5510.5	73.2	1049	4	US-07-757-022B-58
14	5425.5	72.1	1022	4	US-07-757-022B-84
15	5335	70.9	1038	4	US-07-757-022B-74
16	5011	66.6	941	4	US-07-757-022B-114
17	2034.5	27.0	453	4	US-07-757-022B-54
18	1763.5	23.4	423	4	US-07-757-022B-66
19	1754	23.3	422	4	US-07-757-022B-68
20	1460	19.4	372	4	US-07-757-022B-64
21	1188	15.8	220	4	US-07-757-022B-96
22	1136	15.0	209	4	US-07-757-022B-94
23	1121	14.9	208	4	US-07-757-022B-132
24	1111.5	14.8	207	4	US-07-757-022B-116
25	1111.5	14.8	207	4	US-07-757-022B-136
26	1100	14.6	204	4	US-07-757-022B-92
27	1077	14.3	296	4	US-07-757-022B-70

28	1074	14.3	231	4	US-07-757-022B-30	Sequence 30, Appl
29	1041	13.8	192	4	US-07-757-022B-90	Sequence 90, Appl
30	945	12.6	172	4	US-07-757-022B-88	Sequence 88, Appl
31	881	11.7	188	4	US-07-757-022B-32	Sequence 32, Appl
32	871	11.6	156	4	US-07-757-022B-106	Sequence 106, Appl
33	827.5	11.0	157	4	US-07-757-022B-118	Sequence 118, Appl
34	827.5	11.0	157	4	US-07-757-022B-110	Sequence 110, Appl
35	827.5	11.0	157	4	US-07-757-022B-114	Sequence 114, Appl
36	806	10.7	217	4	US-07-757-022B-76	Sequence 76, Appl
37	800	10.6	141	4	US-07-757-022B-80	Sequence 80, Appl
38	750	10.0	131	4	US-07-757-022B-82	Sequence 82, Appl
39	743.5	9.9	237	4	US-07-757-022B-72	Sequence 72, Appl
40	740	9.8	130	4	US-07-757-022B-78	Sequence 78, Appl
41	740	9.8	130	4	US-07-757-022B-86	Sequence 86, Appl
42	706.5	9.4	132	4	US-07-757-022B-140	Sequence 140, Appl
43	702	9.3	8991	4	US-08-714-741-32	Sequence 32, Appl
44	618	8.2	138	4	US-07-757-022B-34	Sequence 34, Appl
45	613	8.1	111	4	US-07-757-022B-28	Sequence 28, Appl

ALIGNMENTS

RESULT 1
US-07-757-022B-2
; Sequence 2, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Geert, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1404 amino acids
; TYPE: AMINO ACID

TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-757-022B-2

Query Match 100.0%; Score 7526; DB 4; Length 1404;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1404; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 MAMKTLPIYLILLISLVFIQOVSSODLSSCAGRCCEGYSRDAITCNCNDVNCCHYMCCDF 60
DB 1 MAMKTLPIYLILLISLVFIQOVSSODLSSCAGRCCEGYSRDAITCNCNDVNCCHYMCCDF 60
QY 61 KRVCIAELSCRCRCESEFERGECDDAOCKKYDCCCDYSEFCLEVNHPSPSSKAP 120
DB 61 KRVCIAELSCRCRCESEFERGECDDAOCKKYDCCCDYSEFCLEVNHPSPSSKAP 120
QY 121 PPSGASQTIKSTTKRSPKPPNKKTKVIESEHIEHSVSENQSSSSSSSSSSSTIW 180
DB 121 PPSGASQTIKSTTKRSPKPPNKKTKVIESEHIEHSVSENQSSSSSSSSSSSTIW 180
QY 181 KIKSSKNAANRELOKLIKVDNKKKRTKKKPTPKRPVVDAGSGLDNGDFKVTTPDST 240
DB 181 KIKSSKNAANRELOKLIKVDNKKKRTKKKPTPKRPVVDAGSGLDNGDFKVTTPDST 240
QY 241 TOHNKVSSTPKITTKAPINPRSLPNSDTSKETSLVYKETTVEKETTITNNKQTSIDG 300
DB 241 TOHNKVSSTPKITTKAPINPRSLPNSDTSKETSLVYKETTVEKETTITNNKQTSIDG 300
QY 301 KEKTSIAKETOSIEKTSADLAPTSXVLAKPTPKAETTTKGPAITTPKEPTTPPKBPAS 360
DB 301 KEKTSIAKETOSIEKTSADLAPTSXVLAKPTPKAETTTKGPAITTPKEPTTPPKBPAS 360
QY 361 TTPKEPTTTIKSAPTTKEBPATTTKSAPTTKEBPATTTKSAPTTKEBPATTTKEP 420
DB 361 TTPKEPTTTIKSAPTTKEBPATTTKSAPTTKEBPATTTKSAPTTKEBPATTTKEP 420
QY 421 APITTKSAPTTKEBPATTTKSAPTTKEBPATTTKSAPTTKEBPATTTKSAPTTKEP 480
DB 421 APITTKSAPTTKEBPATTTKSAPTTKEBPATTTKSAPTTKEBPATTTKSAPTTKEP 480
QY 481 EPAPAPKPPAPPTTKEBPATTTKSAPTTKEBPATTTKSAPTTKEBPATTTKSAPTTKEP 540
DB 481 EPAPAPKPPAPPTTKEBPATTTKSAPTTKEBPATTTKSAPTTKEBPATTTKSAPTTKEP 540
QY 541 TTKAPPTTKEBPATTTKSAPTTKEBPATTTKSAPTTKEBPATTTKSAPTTKEBPATTTKEP 600
DB 541 TTKAPPTTKEBPATTTKSAPTTKEBPATTTKSAPTTKEBPATTTKSAPTTKEBPATTTKEP 600
QY 601 APAPAPKPPAPPTTKEBPATTTKSAPTTKEBPATTTKSAPTTKEBPATTTKSAPTTKEP 660
DB 601 APAPAPKPPAPPTTKEBPATTTKSAPTTKEBPATTTKSAPTTKEBPATTTKSAPTTKEP 660
QY 661 PEEBPATTTKAAAPNTPEKAPPTTKEBPATTTKSAPTTKEBPATTTKSAPTTKEBPATTTKEP 720
DB 661 PEEBPATTTKAAAPNTPEKAPPTTKEBPATTTKSAPTTKEBPATTTKSAPTTKEBPATTTKEP 720
QY 721 APITPKKAPKELAPPTTKEBPATTTKSAPTTKEBPATTTKSAPTTKEBPATTTKSAPTTKEP 780
DB 721 APITPKKAPKELAPPTTKEBPATTTKSAPTTKEBPATTTKSAPTTKEBPATTTKSAPTTKEP 780
QY 781 TAPITLKAPAPPTTKEBPATTTKSAPTTKEBPATTTKSAPTTKEBPATTTKSAPTTKEBPATTTKEP 840
DB 781 TAPITLKAPAPPTTKEBPATTTKSAPTTKEBPATTTKSAPTTKEBPATTTKSAPTTKEBPATTTKEP 840
QY 841 KPAPITPEPTTSEVSTPTTKEPTTIHNSPDSESTPELSABPTPKALENSPKPEGYPT 900
DB 841 KPAPITPEPTTSEVSTPTTKEPTTIHNSPDSESTPELSABPTPKALENSPKPEGYPT 900
QY 901 TKTPAAKPEKMTTAKDKTTERDILTTPETTTAAPKMTKETAATTTKTESKITAATTTQV 960
DB 901 TKTPAAKPEKMTTAKDKTTERDILTTPETTTAAPKMTKETAATTTKTESKITAATTTQV 960
QY 961 TSTTTQDTPPKITTLTKTTTAPKVTTKITITTEINMKKEETAKKPRDRAITNSATTPK 1020

```

```

DB 961 TSTTTQDTPPKITTLTKTTTAPKVTTKITITTEINMKKEETAKKPRDRAITNSATTPK 1020
QY 1021 POKPTAPKPKPTSTYKPKTMRVRKPKTTPPRKMTSTMPBLNPTSRIAEAMLQTTREN 1080
DB 1021 POKPTAPKPKPTSTYKPKTMRVRKPKTTPPRKMTSTMPBLNPTSRIAEAMLQTTREN 1080
QY 1081 QTPNSKLVENPKEDDAGAGETPHMLRPHVMEPTTPMDYLPVPCNGIILNPLS 1140
DB 1081 QTPNSKLVENPKEDDAGAGETPHMLRPHVMEPTTPMDYLPVPCNGIILNPLS 1140
QY 1141 DETNINCKGPDVGLTTLNGLTVAERGHYFWMLSPPSPSPARRITTEWGISPIDVFT 1200
DB 1141 DETNINCKGPDVGLTTLNGLTVAERGHYFWMLSPPSPSPARRITTEWGISPIDVFT 1200
QY 1201 RCNCEGKTFPPKDSQYRFTNDIDAGPKPIFKFGGLTQIYVAALSTAKYKWPESVY 1260
DB 1201 RCNCEGKTFPPKDSQYRFTNDIDAGPKPIFKFGGLTQIYVAALSTAKYKWPESVY 1260
QY 1261 FFKRGGSIOQYIKOEPOKQGRPALNYPVGEKTOVRRRERAIQPSQHTIRQY 1320
DB 1261 FFKRGGSIOQYIKOEPOKQGRPALNYPVGEKTOVRRRERAIQPSQHTIRQY 1320
QY 1321 SPARLAYODKGVLENEKVSILMRGLPNVVTSAISLPIRKPDGYDYAFSKDQYINIDV 1380
DB 1321 SPARLAYODKGVLENEKVSILMRGLPNVVTSAISLPIRKPDGYDYAFSKDQYINIDV 1380
QY 1381 PSRTARAITTRSGQTLKSWYNCP 1404
DB 1381 PSRTARAITTRSGQTLKSWYNCP 1404

RESULT 2
US-07-757-022B-62
; Sequence 62, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Geener, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cseri, Luann

```

REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 876-1170
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 62:
SEQUENCE CHARACTERISTICS:
LENGTH: 1404 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULAR TYPE: protein
US-07-757-0228-62

Query Match 100.0%; Score 7526; DB 4; Length 1404;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1404; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAMKTLPIYLLLSVFIQVSSODLSGCGRGEGSRDATCNCNDYNOHYMECCPDF 60
DB 1 MAMKTLPIYLLLSVFIQVSSODLSGCGRGEGSRDATCNCNDYNOHYMECCPDF 60
QY 61 KAVCTAELSCGKCFESFERGECDDAQCCKTDKCCPDYESTCAEVANPTSPSSKAP 120
DB 61 KAVCTAELSCGKCFESFERGECDDAQCCKTDKCCPDYESTCAEVANPTSPSSKAP 120
QY 121 PPSGASQTIKSTTKSPKPNKKTKVISEEITSEHSVSENOSSSSSSSSSTIM 180
DB 121 PPSGASQTIKSTTKSPKPNKKTKVISEEITSEHSVSENOSSSSSSSSSTIM 180
QY 181 KIKSSKNSANRELQKLLKVKONKQRTKKPTPKPVDVDEAGSGLDNGDFKVTTPDST 240
DB 181 KIKSSKNSANRELQKLLKVKONKQRTKKPTPKPVDVDEAGSGLDNGDFKVTTPDST 240
QY 241 TQHNKVSIPKITTAKPIMPRLPNSDTSKETSITVANKETVEKETTNNKQTSIG 300
DB 241 TQHNKVSIPKITTAKPIMPRLPNSDTSKETSITVANKETVEKETTNNKQTSIG 300
QY 301 KKKTTSAKETOSIEKTSADLAFTSKVLAKPPKATTTTGGPALTTPKEPTTPKEPAS 360
DB 301 KKKTTSAKETOSIEKTSADLAFTSKVLAKPPKATTTTGGPALTTPKEPTTPKEPAS 360
QY 361 TTPKEPTPTTIKSAPTTPKEPAATTTKSAPTTPKEPAATTTTPKEPAATTTKEP 420
DB 361 TTPKEPTPTTIKSAPTTPKEPAATTTKSAPTTPKEPAATTTTPKEPAATTTKEP 420
QY 421 APTTTSAPTTPEKAPATTPPKAPATTPKEPAATTPPTTPKEPAATTPKEP 480
DB 421 APTTTSAPTTPEKAPATTPPKAPATTPKEPAATTPPTTPKEPAATTPKEP 480
QY 481 BPAPTPKAPATTPKEPAATTPKEPAATTTKEPSPTTKEPAATTTTKEPAP 540
DB 481 BPAPTPKAPATTPKEPAATTPKEPAATTTKEPSPTTKEPAATTTTKEPAP 540
QY 541 TTKSAPTTPEKSPPTTKEPAATTPKEPAATTPPKAPATTPKEPAATTTTKEP 600
DB 541 TTKSAPTTPEKSPPTTKEPAATTPKEPAATTPPKAPATTPKEPAATTTTKEP 600
QY 601 APTAPKEPAATTPKEPAATTPPKLPTTPEKLAFTTPEKAPATTPPEELAP 660
DB 601 APTAPKEPAATTPKEPAATTPPKLPTTPEKLAFTTPEKAPATTPPEELAP 660
QY 661 PSEBAPTTKAAABNTPEKAPATTPKEPAATTPKEPAATTTTKEPAATTTKEP 720
DB 661 PSEBAPTTKAAABNTPEKAPATTPKEPAATTPKEPAATTTTKEPAATTTKEP 720
QY 721 APTTPKAPAKELAPTTTKEPTSTSDKAPATTPKSTAPATTPKEPAATTPK 780
DB 721 APTTPKAPAKELAPTTTKEPTSTSDKAPATTPKSTAPATTPKEPAATTPK 780
QY 781 TAPTTTKEPAATTPKAPAKELAPTTTKEPTSTSDKAPATTPKSTAPATTPK 840
DB 781 TAPTTTKEPAATTPKAPAKELAPTTTKEPTSTSDKAPATTPKSTAPATTPK 840

QY 841 KPAPTTPEPTSPPTSEVSTPTTKEPTTIHKSPODESTPELSAPPPKALENSPKRGVPT 900
DB 841 KPAPTTPEPTSPPTSEVSTPTTKEPTTIHKSPODESTPELSAPPPKALENSPKRGVPT 900
QY 901 TKTPAATKEMTTTAKDKTERDLRTTPETTTAAKMTKETATTTTEKTESKITATTTQV 960
DB 901 TKTPAATKEMTTTAKDKTERDLRTTPETTTAAKMTKETATTTTEKTESKITATTTQV 960
QY 961 TSTTODTTPFKITTLTKTTTLAPKVTYTKKITTTTEINANKPEETAKPORAATNSKATTPK 1020
DB 961 TSTTODTTPFKITTLTKTTTLAPKVTYTKKITTTTEINANKPEETAKPORAATNSKATTPK 1020
QY 1021 POKPTKAPKPTSTTKKPKTPMVRKPKTPPTPRKMTSTMPBLNPTSRISAMLOTTTRPN 1080
DB 1021 POKPTKAPKPTSTTKKPKTPMVRKPKTPPTPRKMTSTMPBLNPTSRISAMLOTTTRPN 1080
QY 1081 QTPNSKLVENPKSBDAGAGEBTPHMLRPHVFMPEVTPDMDYLPVBNQIINPMLS 1140
DB 1081 QTPNSKLVENPKSBDAGAGEBTPHMLRPHVFMPEVTPDMDYLPVBNQIINPMLS 1140
QY 1141 DETNLCNGKPVUGLTTLNKGTLVAFRGHYFMMLSPSPSPAPARTTEWGISPIDVTPT 1200
DB 1141 DETNLCNGKPVUGLTTLNKGTLVAFRGHYFMMLSPSPSPAPARTTEWGISPIDVTPT 1200
QY 1201 RCNCGKPTFFKDSQYMRPTNDIKDAGYKPKIFKPGGLTGOIYVALSTAKYKNPESVY 1260
DB 1201 RCNCGKPTFFKDSQYMRPTNDIKDAGYKPKIFKPGGLTGOIYVALSTAKYKNPESVY 1260
QY 1261 FFKGGSIIQYIYKQEPVQKCPGRPALNYPYGEMTQVRRRRFERALGPSQTHIRIQY 1320
DB 1261 FFKGGSIIQYIYKQEPVQKCPGRPALNYPYGEMTQVRRRRFERALGPSQTHIRIQY 1320
QY 1321 SPARLAYODKGVLANHEVKVSIIMRGLPNVVTSAISLPINRKGDGYDAFSDQYNNIDV 1380
DB 1321 SPARLAYODKGVLANHEVKVSIIMRGLPNVVTSAISLPINRKGDGYDAFSDQYNNIDV 1380
QY 1381 PSRTARITTSRGOTLSKVNCP 1404
DB 1381 PSRTARITTSRGOTLSKVNCP 1404

RESULT 3
US-07-757-0228-40
Sequence 40, Application US/077570228
Patent No. 6433142
GENERAL INFORMATION:
APPLICANT: Geener, Thomas G.
APPLICANT: Clark, Stephen C.
APPLICANT: Turner, Katherine
APPLICANT: Hewick, Rodney M.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESSES:
ADDRESSER: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/757,0228
FILING DATE: 19910910
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/546,114

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/757,022B
FILING DATE: 19910910
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cbeir, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 52:
SEQUENCE CHARACTERISTICS:
LENGTH: 1363 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-757-022B-52

Query Match 96.3%; Score 7245.5; DB 4; Length 1363;

Best Local Similarity 97.1%; Pred. No. 0;

Matches 1363; Conservative 0; Mismatches 0; Indels 41; Gaps 1;

QY 1 MAMKTLPIYLLLSVFIQVSSODLSGCAGRCGEYSRDATCNCYNQHYMECCPDE 60
DB 1 MAMKTLPIYLLLSVFIQVSSODLSGCAGRCGEYSRDATCNCYNQHYMECCPDE 25
QY 61 KAVCTAELSCGKCFSPFGRGRCDDAQCCKTDKCCPDYESTCAVHNPTSPSSKAP 120
DB 26 -----ELSCGKCFSPFGRGRCDDAQCCKTDKCCPDYESTCAVHNPTSPSSKAP 79
QY 121 PPSGASQTIKSTTKRSPKPNKKTKKVTSEBEITEHSVENQSSSSSSSSSSSTIM 180
DB 80 PPSGASQTIKSTTKRSPKPNKKTKKVTSEBEITEHSVENQSSSSSSSSSSSTIM 139
QY 181 KIKSSKNSAANRELQKKLVKDNKNKRTKKKPTPKPPVNDGASGLDNGDFKVTEDTST 240
DB 140 KIKSSKNSAANRELQKKLVKDNKNKRTKKKPTPKPPVNDGASGLDNGDFKVTEDTST 199
QY 241 TOHNKVTSPBKITTAAPINRPSLPNSDTSKETSITLUNKETTVETKTTTNNKQSTDG 300
DB 200 TOHNKVTSPBKITTAAPINRPSLPNSDTSKETSITLUNKETTVETKTTTNNKQSTDG 259
QY 301 KKKTSKAKETOSIEKTSANDLAPTSKVLAKPTKATTTTGGALTTPKKPTPTPKEPAS 360
DB 260 KKKTSKAKETOSIEKTSANDLAPTSKVLAKPTKATTTTGGALTTPKKPTPTPKEPAS 319
QY 361 TTPKKEPTTIKSAPTTPKEPAFTTTKSAFTTPKEPAFTTTKKEPAFTTTKKEPA 420
DB 320 TTPKKEPTTIKSAPTTPKEPAFTTTKSAFTTPKEPAFTTTKKEPAFTTTKKEPA 379
QY 421 APTTTSAPTTPEKPAFTTPKKAFTTPKPAFTTPKEPAFTTPKEPAFTTTKKEPA 480
DB 380 APTTTSAPTTPEKPAFTTPKKAFTTPKPAFTTPKEPAFTTPKEPAFTTTKKEPA 439
QY 481 BPAATPKKPAFTTPKEPAFTTPKEPAFTTTKSPSTTPKEPAFTTTKSAFTTTKKEPA 540
DB 440 BPAATPKKPAFTTPKEPAFTTPKEPAFTTTKSPSTTPKEPAFTTTKSAFTTTKKEPA 499
QY 541 TTKSAPTTPEKSPSTTTKKEPAFTTPKEPAFTTTKKAFTTTKKEPAFTTTKKEPA 600
DB 541 TTKSAPTTPEKSPSTTTKKEPAFTTPKEPAFTTTKKAFTTTKKEPAFTTTKKEPA 600

DB 500 TTKSAPTTPEKSPSTTTKKEPAFTTPKEPAFTTTKKAFTTTKKEPAFTTTKKEPA 559
QY 601 APTAPKEPAFTTPKEPAFTTTKKAFTTTKKEPAFTTTKKEPAFTTTKKEPAFTTT 660
DB 560 APTAPKEPAFTTTKKEPAFTTTKKAFTTTKKEPAFTTTKKEPAFTTTKKEPAFTTT 619
QY 661 PEEPAFTTTKKAAPNTPKKPAFTTPKEPAFTTTKKEPAFTTTKKEPAFTTTKKEPA 720
DB 620 PEEPAFTTTKKAAPNTPKKPAFTTPKEPAFTTTKKEPAFTTTKKEPAFTTTKKEPA 679
QY 721 APTPKKPAFTTTKKEPAFTTTKKEPAFTTTKKEPAFTTTKKEPAFTTTKKEPAFT 780
DB 680 APTPKKPAFTTTKKEPAFTTTKKEPAFTTTKKEPAFTTTKKEPAFTTTKKEPAFT 739
QY 781 TAPTTKKEPAFTTTKKEPAFTTTKKEPAFTTTKKEPAFTTTKKEPAFTTTKKEPA 840
DB 740 TAPTTKKEPAFTTTKKEPAFTTTKKEPAFTTTKKEPAFTTTKKEPAFTTTKKEPA 799
QY 841 KPAFTTPEPTTSEVSTPTTKEBPTTIHKSPDESTPELSAAPTKEALNSPKGQVPT 900
DB 800 KPAFTTPEPTTSEVSTPTTKEBPTTIHKSPDESTPELSAAPTKEALNSPKGQVPT 859
QY 901 TKTPAATKEEMTTAANDKTTERRDLRTTPTTAAAPMTKETAATTEKTESKITATTTQV 960
DB 860 TKTPAATKEEMTTAANDKTTERRDLRTTPTTAAAPMTKETAATTEKTESKITATTTQV 919
QY 961 TSTTTODTTPFKITTLTKITTLAPKVTYTTTKITTEINMKPSETAKPKRATNSKATTK 1020
DB 920 TSTTTODTTPFKITTLTKITTLAPKVTYTTTKITTEINMKPSETAKPKRATNSKATTK 979
QY 1021 POKPTKAPKKPTSTKKPKTTPRVKPKTTPTRKMTSTMPBLNPTSRISAEAMQTTTRN 1080
DB 980 POKPTKAPKKPTSTKKPKTTPRVKPKTTPTRKMTSTMPBLNPTSRISAEAMQTTTRN 1039
QY 1081 QTPNSKLVENPKSEBAGABETPHMLLRPHVFMPEVTPDMVDYLPRVNOGIIINPMLS 1140
DB 1040 QTPNSKLVENPKSEBAGABETPHMLLRPHVFMPEVTPDMVDYLPRVNOGIIINPMLS 1099
QY 1141 DETNTCNGKPVNGLTTLRNGTLVAFRGHFMMLSPSPSPSPARRITVEWGISPTVTPT 1200
DB 1100 DETNTCNGKPVNGLTTLRNGTLVAFRGHFMMLSPSPSPSPARRITVEWGISPTVTPT 1159
QY 1201 RCNCEKTPFPFQDSQVWFRTNDIKDAGVKKPIFKGFGGLTGOIIVAALSTAKKNMPESEY 1260
DB 1160 RCNCEKTPFPFQDSQVWFRTNDIKDAGVKKPIFKGFGGLTGOIIVAALSTAKKNMPESEY 1219
QY 1261 PFKGGSIOQYIYKQEPVQKCPGRPALNYPYVGEMTQVRRRRFERALGPSQTHITRIQY 1320
DB 1220 PFKGGSIOQYIYKQEPVQKCPGRPALNYPYVGEMTQVRRRRFERALGPSQTHITRIQY 1279
QY 1321 SPARLAYQDKVLANHEVKVSIIMRGLPNVVTSAISLPIIRKPDGIDYVAFSKDQYINIDV 1380
DB 1280 SPARLAYQDKVLANHEVKVSIIMRGLPNVVTSAISLPIIRKPDGIDYVAFSKDQYINIDV 1339
QY 1381 PSRTARAITTRSGQTLISKVWVNC 1404
DB 1340 PSRTARAITTRSGQTLISKVWVNC 1363

RESULT 5
US-07-757-022B-48
Sequence 48. Application US/07757022B
Patent No. 6433142
GENERAL INFORMATION:
APPLICANT: Gesner, Thomas G.
APPLICANT: Clark, Stephen C.
APPLICANT: Turner, Katherine
APPLICANT: Hewick, Rodney M.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive

CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/757,022B
FILING DATE: 19910910
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cseert, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 876-5851
TELEPHONE: (617) 876-1170
INFORMATION FOR SEQ ID NO: 48:
SEQUENCE CHARACTERISTICS:
LENGTH: 1354 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-757-022B-48

Query Match 96.1%; Score 7232; DB 4; Length 1354;
Best Local Similarity 96.4%; Pred. No. 0;
Matches 1354; Conservative 0; Mismatches 0; Indels 50; Gaps 1;
QY 1 MAMKTLPIYLLLSVFVIOQVSSODLSSCAGRCGEGRSDATCNCDCVNCQHMECCDF 60
DB 1 MAMKTLPIYLLLSVFVIOQVSSODLSSCAGRCGEGRSDATCNCDCVNCQHMECCDF 60
QY 61 KRVCTAELSCGRCPESFERGECDDAQCKKYDKCCPDYESFCAEVNPTSPSSKAP 120
DB 61 KRVCTAELSCGRCPESFERGECDDAQCKKYDKCCPDYESFCAEVNPTSPSSKAP 120
QY 121 PPSGASQTIKSTTKRSPKPKKKTKKVI ESEBITHEHSVSENDESSSSSSSSSSSTIW 180
DB 106 -----EHSVSENDESSSSSSSSSSSSSTIW 130
QY 181 KIKSSKSAARELOKLVKNDKKNRTKKKPTKPPVVDAGSLDNGDKVTTPDST 240
DB 131 KIKSSKSAARELOKLVKNDKKNRTKKKPTKPPVVDAGSLDNGDKVTTPDST 190
QY 241 TQHNKSVSPKITTKAPINPRPSLPNSDTSKETSLVANKETVETKETTNNKQSTDG 300
DB 191 TQHNKSVSPKITTKAPINPRPSLPNSDTSKETSLVANKETVETKETTNNKQSTDG 250
QY 301 KEKTTSAKETOSIEKTSADIAPTS KVLAKPTPKAETTTKGPALITTPKEPTTPKEPAS 360
DB 251 KEKTTSAKETOSIEKTSADIAPTS KVLAKPTPKAETTTKGPALITTPKEPTTPKEPAS 310
QY 361 TTPKEPTTPTKSAPTTKEBPATTTKSAPTTKEBPATTTKEBPATTTKEBPATTTKEP 420
DB 311 TTPKEPTTPTKSAPTTKEBPATTTKSAPTTKEBPATTTKEBPATTTKEBPATTTKEP 370
QY 421 APTTTSAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPPTTPTPKEBPATTTKEBPATTPK 480

DB 371 APTTTSAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPPTTPTPKEBPATTTKEBPATTPK 430
QY 481 EPAPTPAKKAPATTPKEBPATTTKEBPATTTKEBPATTTKEBPATTTKSAPTTTKEBPAT 540
DB 431 EPAPTPAKKAPATTPKEBPATTTKEBPATTTKEBPATTTKEBPATTTKSAPTTTKEBPAT 490
QY 541 TTKSAPTTKEBPATTTKEBPATTTKEBPATTTPKBPAPTTPKBPAPTTKEBPATTTKEBP 600
DB 491 TTKSAPTTKEBPATTTKEBPATTTKEBPATTTPKBPAPTTPKBPAPTTKEBPATTTKEBP 550
QY 601 APTAPKEBPATTTKEBPATTTKEBPATTTKEBPATTTKEBPATTTKEBPATTTKEBPATTT 660
DB 551 APTAPKEBPATTTKEBPATTTKEBPATTTKEBPATTTKEBPATTTKEBPATTTKEBPATTT 610
QY 661 PEEBPATTPKAAANTPEKBPATTTKEBPATTTKEBPATTTKEBPATTTKEBPATTTKEBP 720
DB 611 PEEBPATTPKAAANTPEKBPATTTKEBPATTTKEBPATTTKEBPATTTKEBPATTTKEBP 670
QY 721 APTTPKKAPAPKELAPTTTKEPTSTSDKAPATTTKEBPATTTKEBPATTTKEBPATTTKEP 780
DB 671 APTTPKKAPAPKELAPTTTKEPTSTSDKAPATTTKEBPATTTKEBPATTTKEBPATTTKEP 730
QY 781 TAPTTKEBPATTTKEBPATTTKEBPATTTKEBPATTTKEBPATTTKEBPATTTKEBPATTT 840
DB 731 TAPTTKEBPATTTKEBPATTTKEBPATTTKEBPATTTKEBPATTTKEBPATTTKEBPATTT 790
QY 841 KPAATTPETPTTSEVSTPTTKEPTTIHKSDEESTPELSABPTPALENSPEKPGVPT 900
DB 791 KPAATTPETPTTSEVSTPTTKEPTTIHKSDEESTPELSABPTPALENSPEKPGVPT 850
QY 901 TKTPAATPEMTTAKOTTERDRTTPETTAARPKMTKEATTTTEKTESKIRATTTQV 960
DB 851 TKTPAATPEMTTAKOTTERDRTTPETTAARPKMTKEATTTTEKTESKIRATTTQV 910
QY 961 TSTTTODTTPKLTITLTKITTLAPAVTTTKUITITTEIMNKEETAKDRATNSKATTPK 1020
DB 911 TSTTTODTTPKLTITLTKITTLAPAVTTTKUITITTEIMNKEETAKDRATNSKATTPK 970
QY 1021 POKPTKAPKKTSTKCKRTMPRVKPKTTPPRKMTSTMPRLNTSKIAEAMLOTTTRPN 1080
DB 971 POKPTKAPKKTSTKCKRTMPRVKPKTTPPRKMTSTMPRLNTSKIAEAMLOTTTRPN 1030
QY 1081 QTPNSKIVEVNPKEDEGAGETPHMLRPHVMPETTPMDYLPRVPOGIIINPMLS 1140
DB 1031 QTPNSKIVEVNPKEDEGAGETPHMLRPHVMPETTPMDYLPRVPOGIIINPMLS 1090
QY 1141 DETNICNGKVPDGLTTLRNGTLVAFRGHYFWMLSPPSPSPARRITEVWGIRSPIDTVFT 1200
DB 1091 DETNICNGKVPDGLTTLRNGTLVAFRGHYFWMLSPPSPSPARRITEVWGIRSPIDTVFT 1150
QY 1201 RCNCEGKTPFFKQSVYRFTNDIDAGYPKPIFFGFGGLGQIYVAIASTAYKXWPSVY 1260
DB 1151 RCNCEGKTPFFKQSVYRFTNDIDAGYPKPIFFGFGGLGQIYVAIASTAYKXWPSVY 1210
QY 1261 FFKRGSISQYIYKQEBVOKOPGRPALANYPVGEEMQVRRRRERAIIGSQHTTIRIQY 1320
DB 1211 FFKRGSISQYIYKQEBVOKOPGRPALANYPVGEEMQVRRRRERAIIGSQHTTIRIQY 1270
QY 1321 SPARLAQODKGVLANEVKSLMRGLPNVYTSALSNIRKPDGYDYVYASKQDYNNIDV 1380
DB 1271 SPARLAQODKGVLANEVKSLMRGLPNVYTSALSNIRKPDGYDYVYASKQDYNNIDV 1330
QY 1381 PSRTARAITTRSGQTLKRWYNCP 1404
DB 1331 PSRTARAITTRSGQTLKRWYNCP 1354

RESULT 6
US-07-757-022B-46
; Sequence 46, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:

```

/ APPLICANT: Geener, Thomas G.
/ APPLICANT: Clark, Stephen C.
/ APPLICANT: Turner, Katherine
/ APPLICANT: Hewick, Rodney M.
/ TITLE OF INVENTION: Megakaryocyte Stimulating Factors
/ NUMBER OF SEQUENCES: 143
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Genetics Institute, Inc.
/ STREET: 87 Cambridgepark Drive
/ CITY: Cambridge
/ STATE: Massachusetts
/ COUNTRY: U.S.A.
/ ZIP: 02140
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patent Release #1.0, Version #1.25
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/07/757,022B
/ FILING DATE: 19910910
/ CLASSIFICATION: 530
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 07/643,502
/ FILING DATE: 18-JAN-1991
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 07/546,114
/ FILING DATE: 29-JUN-1990
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 07/457,196
/ FILING DATE: 29-DEC-1989
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 07/390,901
/ FILING DATE: 08-AUG-1989
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Geier, Luann
/ REGISTRATION NUMBER: 31,822
/ REFERENCE/DOCKET NUMBER: GI 5190
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (617) 876-1170
/ TELEFAX: (617) 876-5851
/ INFORMATION FOR SEQ ID NO: 46:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 1320 amino acids
/ TYPE: AMINO ACID
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/ US-07-757-022B-46

```

Query Match 93.2%; Score 7011; DB 4; Length 1320;
Best Local Similarity 94.0%; Pired. No. 0;
Matches 1340; Conservative 0; Mismatches 0; Indels 84; Gaps 2;

```

QY 1 MAMKTLPIYLLLSLVFVIQOVSSQDLSACGRGEGYRDATCNCDCYNQHYMECCPDF 60
DB 1 MAMKTLPIYLLLSLVFVIQOVSSQ----- 25
QY 61 KRVCTALISCKGRGCFSPERGRRCDCDAOCKKYDKCCPYESFCAEVHNPTSPSSKAP 120
DB 26 -----ELSCGKGRGCFSPERGRRCDCDAOCKKYDKCCPYESFCAEVHNPTSPSSKAP 79
QY 121 PPSGASQTIKSTTKRSPKPKPNKKTKKVI ESEETIEHSEVSEHQESSSSSSSSSSSTIM 180
DB 80 PPSGASQTIKSTTKRSPKPKPNKKTKKVI ESEETIE----- 115
QY 181 KIKSRSKNSANRELQCKLKVKONKKNRTKKKPTPKPPVVDGASGLDNGDFKVTTPDST 240
DB 116 -----VKONKKNRTKKKPTPKPPVVDGASGLDNGDFKVTTPDST 156
QY 241 TQHNKVSISPKITTAQINPRSELPPNSDTSKETSITVNKETTVEETKTTTNKQTSIDG 300
DB 157 TQHNKVSISPKITTAQINPRSELPPNSDTSKETSITVNKETTVEETKTTTNKQTSIDG 216

```

```

QY 301 KERTTSKETSOSIEKTSASAKLAPTSKVLAKPTPKAETTTKGPALTTPKEPTPTPKSPAS 360
DB 217 KERTTSKETSOSIEKTSASAKLAPTSKVLAKPTPKAETTTKGPALTTPKEPTPTPKSPAS 276
QY 361 TTPKEPTPTTIKSAAPTTKBPAPTTTKSAAPTTKBPAPTTTKBPAPTTTKBPAPTTTKBP 420
DB 277 TTPKEPTPTTIKSAAPTTKBPAPTTTKSAAPTTKBPAPTTTKBPAPTTTKBPAPTTTKBP 336
QY 421 APTTTKSAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAP 480
DB 337 APTTTKSAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAP 396
QY 481 BPAPTAAPKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBP 540
DB 397 BPAPTAAPKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBP 456
QY 541 TTKSAAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBP 600
DB 457 TTKSAAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBP 516
QY 601 APTAPKEBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBP 660
DB 517 APTAPKEBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBP 576
QY 661 PEPAPPTPKAAAPNTPKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBP 720
DB 577 PEPAPPTPKAAAPNTPKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBP 636
QY 721 APTPKKAPKELAPTTKBPSTTSDDKAPPTPKGTAAPTTKBPAPTTKBPAPTTKBPAPTTKBP 780
DB 637 APTPKKAPKELAPTTKBPSTTSDDKAPPTPKGTAAPTTKBPAPTTKBPAPTTKBPAPTTKBP 696
QY 781 TAPTLKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAP 840
DB 697 TAPTLKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAP 756
QY 841 KPAPTTPEPTPTTSVSTPTTKBPSTTHKSPDSTPELSAPTKALENSPEKSGVPT 900
DB 757 KPAPTTPEPTPTTSVSTPTTKBPSTTHKSPDSTPELSAPTKALENSPEKSGVPT 816
QY 901 TKTPAATKEMTTAKDKITERDLRTTBETTTAAKMTKETATTTKETSKITATTTTQV 960
DB 817 TKTPAATKEMTTAKDKITERDLRTTBETTTAAKMTKETATTTKETSKITATTTTQV 876
QY 961 TSTTQDTTPFKITTLKTTTLAPKVTYTTKITTTEINMKPEETAKKORATNSKATTPK 1020
DB 877 TSTTQDTTPFKITTLKTTTLAPKVTYTTKITTTEINMKPEETAKKORATNSKATTPK 936
QY 1021 POKPTKAPKKPTSTKKPKTPMVRKPKTTPPRKMTSTMBELNPTSRISAEMLQTTTRN 1080
DB 937 POKPTKAPKKPTSTKKPKTPMVRKPKTTPPRKMTSTMBELNPTSRISAEMLQTTTRN 996
QY 1081 QTPNSKLVENVNKSBDAGABETPHMLLRPHVFEVLTDMQDYLRPVNQCIIINPMLS 1140
DB 997 QTPNSKLVENVNKSBDAGABETPHMLLRPHVFEVLTDMQDYLRPVNQCIIINPMLS 1056
QY 1141 DETNINCKGKVDGLTTLNGLTVAFRGHYFMMLSPSPSPSPARRITTEWGISPIDTYVT 1200
DB 1057 DETNINCKGKVDGLTTLNGLTVAFRGHYFMMLSPSPSPSPARRITTEWGISPIDTYVT 1116
QY 1201 RCNCGKTPFPKDSQYMFNDIKDAGYPKPIFKGFGGLTGOIVAAALSTAKTKMPESEY 1260
DB 1117 RCNCGKTPFPKDSQYMFNDIKDAGYPKPIFKGFGGLTGOIVAAALSTAKTKMPESEY 1176
QY 1261 PFKRGSIQOYLYKQEPVQKCGRRPALNTPYVGEMTOYRRRRFBRALGPSQTHIRIQY 1320
DB 1177 PFKRGSIQOYLYKQEPVQKCGRRPALNTPYVGEMTOYRRRRFBRALGPSQTHIRIQY 1236
QY 1321 SPARLAIQDKGLVLAHEVYSILMRGLPNVVTSAISLPINRKDGVDYIAFSKQOYNNIDV 1380
DB 1237 SPARLAIQDKGLVLAHEVYSILMRGLPNVVTSAISLPINRKDGVDYIAFSKQOYNNIDV 1296
QY 1381 PERTARAITTRGQTLKVMVNCPL 1404

```


Db 1297 PSRTBARITTRSGLSKVWNCIP 13320

|||||

RESULT 7

US-07-757-022B-60

; Sequence 60, Application US/07757022B

; Patent No. 6433142

GENERAL INFORMATION:

APPLICANT: Gessner, Thomas G.

APPLICANT: Clark, Stephen C.

APPLICANT: Turner, Katherine

APPLICANT: Hewick, Rodney M.

TITLE OF INVENTION: Megakaryocyte Stimulating Factors

NUMBER OF SEQUENCES: 143

CORRESPONDENCE ADDRESS:

ADDRESSEE: Genetics Institute, Inc.

STREET: 87 Cambridgepark Drive

CITY: Cambridge

STATE: Massachusetts

COUNTRY: U.S.A.

ZIP: 02140

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/757,022B

FILING DATE: 19910910

CLASSIFICATION: 530

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/643,502

FILING DATE: 18-JAN-1991

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/546,114

FILING DATE: 29-JUN-1990

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/457,196

FILING DATE: 29-DEC-1989

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/390,901

FILING DATE: 08-AUG-1989

ATTORNEY/AGENT INFORMATION:

NAME: Cseer, Luann

REGISTRATION NUMBER: 31,822

REFERENCE/DOCKET NUMBER: GI 5190

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617)876-1170

TELEFAX: (617)876-5851

INFORMATION FOR SEQ ID NO: 60:

SEQUENCE CHARACTERISTICS:

LENGTH: 1320 amino acids

TYPE: AMINO ACID

TOPOLOGY: linear

MOLECULE TYPE: protein

US-07-757-022B-60

Query Match 93.2%; Score 7011; DB 4; Length 1320;

Best Local Similarity 94.0%; Pred. No. 0;

Matches 1320; Conservative 0; Mismatches 0; Indels 84; Gaps 2;

QY 1 MAMTLPYLLLLSVFIQVSSQ----- 25

Db 1 MAMTLPYLLLLSVFIQVSSQ----- 25

QY 61 KRVCTAELISGRCFESPERGECDDAQCCKYDKCCPDYESFCAEVNPTSPSSSKAP 120

Db 26 -----ELSCGRCFESPERGECDDAQCCKYDKCCPDYESFCAEVNPTSPSSSKAP 79

QY 121 PPSGASQTIKSTTRSPKPKTKKVIIESEITEHSVSENOESSSSSSSSSSSTIW 180

Db 80 PPSGASQTIKSTTRSPKPKTKKVIIESEITE----- 115

QY 181 KIKSSKNSAANRELQKLVKONKKNRKKKPKPPVVDAGSLDNGDKTTPDTST 240

Db 116 -----VKONKKNRKKKPKPPVVDAGSLDNGDKTTPDTST 156

QY 241 TQHNKSVSPKITTAKPINRPSLPPNSDTSKETSLTVNKEETVETKETTNNKQTSIDG 300

Db 157 TQHNKSVSPKITTAKPINRPSLPPNSDTSKETSLTVNKEETVETKETTNNKQTSIDG 216

QY 301 KEKTTSAKETOSIKTSKADLAPTSKVLAKPTPAETTTKGPALTTPKEPTTPKEPAS 360

Db 217 KEKTTSAKETOSIKTSKADLAPTSKVLAKPTPAETTTKGPALTTPKEPTTPKEPAS 276

QY 361 TTPKEPTTPKSAPTTPKEBAPTTPKSAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKE 420

Db 277 TTPKEPTTPKSAPTTPKEBAPTTPKSAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKE 336

QY 421 APTTKSAPTPKBPAPTPPKBPAPTPPKBPAPTPPKBPAPTPPKBPAPTPPKBPAPTP 480

Db 337 APTTKSAPTPKBPAPTPPKBPAPTPPKBPAPTPPKBPAPTPPKBPAPTPPKBPAPTP 396

QY 481 EPAPTAPKBPAPTPPKBPAPTPPKBPAPTPPKBPAPTPPKBPAPTPPKBPAPTP 540

Db 397 EPAPTAPKBPAPTPPKBPAPTPPKBPAPTPPKBPAPTPPKBPAPTPPKBPAPTP 456

QY 541 TTKSAPTPKBPAPTPPKBPAPTPPKBPAPTPPKBPAPTPPKBPAPTPPKBPAPTP 600

Db 457 TTKSAPTPKBPAPTPPKBPAPTPPKBPAPTPPKBPAPTPPKBPAPTPPKBPAPTP 516

QY 601 APTAPKBPAPTPPKBPAPTPPKBPAPTPPKBPAPTPPKBPAPTPPKBPAPTP 660

Db 517 APTAPKBPAPTPPKBPAPTPPKBPAPTPPKBPAPTPPKBPAPTPPKBPAPTP 576

QY 661 PEBPAPTPPKAAANTKEBPAPTPPKBPAPTPPKBPAPTPPKBPAPTPPKBPAPTP 720

Db 577 PEBPAPTPPKAAANTKEBPAPTPPKBPAPTPPKBPAPTPPKBPAPTPPKBPAPTP 636

QY 721 APTPKBPAPKELAPTTKBPSTSTSDKAPATTKGAPATTKGAPATTKGAPATTKG 780

Db 637 APTPKBPAPKELAPTTKBPSTSTSDKAPATTKGAPATTKGAPATTKGAPATTKG 696

QY 781 TAPTLKEBPAPTPPKBPAPKELAPTTKGPSTSDAPAPTPPKBPAPTPPKBPAPTP 840

Db 697 TAPTLKEBPAPTPPKBPAPKELAPTTKGPSTSDAPAPTPPKBPAPTPPKBPAPTP 756

QY 841 KPAPTTPEPTPTTSVSTPTTKETTHKSPBESITPELSAETPKALENSPREGVPT 900

Db 757 KPAPTTPEPTPTTSVSTPTTKETTHKSPBESITPELSAETPKALENSPREGVPT 816

QY 901 TKTPAATKPEMTTAKKTERDRTTPETTAPAKMTKETATTEKTESKITATTTQV 960

Db 817 TKTPAATKPEMTTAKKTERDRTTPETTAPAKMTKETATTEKTESKITATTTQV 876

QY 961 TSTTTQDTPEPKITLTKTTLLAPKVTTTKKIIITTEIMNKPEETAKPKDRATNSKATTP 1020

Db 877 TSTTTQDTPEPKITLTKTTLLAPKVTTTKKIIITTEIMNKPEETAKPKDRATNSKATTP 936

QY 1021 POKPTAPPKPTSTKCKTMRVAKKPTTPPKRMTSTMBELNPTSIAEAMLOTTTRPN 1080

Db 937 POKPTAPPKPTSTKCKTMRVAKKPTTPPKRMTSTMBELNPTSIAEAMLOTTTRPN 996

QY 1081 QTPNSKLVENPSESDGAGBETPHMLRPHVMPVETDMDVLPRVPOGIIINMPLS 1140

Db 997 QTPNSKLVENPSESDGAGBETPHMLRPHVMPVETDMDVLPRVPOGIIINMPLS 1056

QY 1141 DETNINCGKPVDDGLITLRLNGTLVAFRGHYFWMLSPPSPSPPARITVWGIPSDITVFT 1200

Db 1057 DETNINCGKPVDDGLITLRLNGTLVAFRGHYFWMLSPPSPSPPARITVWGIPSDITVFT 1116

QY 1201 RNCCEGKTPFPKQSVYRFTNDIKDAGYKPIIFKGFGGLTGQIYALSTAKYKWPSPSVY 1260

Db 1117 RNCCEGKTPFPKQSVYRFTNDIKDAGYKPIIFKGFGGLTGQIYALSTAKYKWPSPSVY 1176

QY 1261 FFRGGSIQQYIYKQEPVQKCPGRPALNYPVGEVTVRRRRFERAIGPSQTHIRIQY 1320
DB 1177 FFRGGSIQQYIYKQEPVQKCPGRPALNYPVGEVTVRRRRFERAIGPSQTHIRIQY 1236
QY 1321 SPARLAVQDGVVLANEKKVSIILWGLPNTVTSALSLPNIKKPGDYVYAFSSKOYNNIDY 1380
DB 1237 SPARLAVQDGVVLANEKKVSIILWGLPNTVTSALSLPNIKKPGDYVYAFSSKOYNNIDY 1296
QY 1381 PSRTARATTSRGOTLSKVWYNCP 1404
DB 1297 PSRTARATTSRGOTLSKVWYNCP 1320

RESULT 8

US-07-757-022B-42
Sequence 42, Application US/07757022B
Patent No. 6433142
GENERAL INFORMATION:
APPLICANT: Geener, Thomas G.
APPLICANT: Clark, Stephen C.
APPLICANT: Turner, Katherine
APPLICANT: Hewick, Rodney M.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSER: Genetics Institute, Inc.
STREET: 87 Cambridge Park Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/757,022B
FILING DATE: 19910910
CLASSIFICATION: 530
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Geier, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 42:
SEQUENCE CHARACTERISTICS:
LENGTH: 1311 amino acids
TYPE: AMINO ACID
TOPOLOGY: 1linear
MOLECULAR TYPE: Protein
US-07-757-022B-42

QY Query Match 93.1%; Score 7007.5; DB 4; Length 1311;
Best Local Similarity 93.4%; Pred. No. 0;
Matches 1311; Conservative 0; Mismatches 0; Indels 93; Gaps 1;
QY 1 MAMKTLPIYLILLLSVFIQVVSQDLSGACGCEGYSRDATCNCNDVNCQHYMECCPDF 60
|||||

DB 1 MAMKTLPIYLILLLSVFIQVVSQDLSGACGCEGYSRDATCNCNDVNCQHYMECCPDF 60
QY 61 KAVCTIAELSCRCRESFERGECCTDQCKKTYDKCCPDYSEFCABVHNPTSPSSKAP 120
DB 61 KAVCTIAELSCRCRESFERGECCTDQCKKTYDKCCPDYSEFCABVHNPTSPSSKAP 106
QY 121 PSAGASQTIKSTKSSPKPKNKKTKYVLESBEIRIEHVSANQSSSSSSSSSTIIV 180
DB 107 ----- 106
QY 181 KIKSSKNSAANRELQKLVKDNKNKRTKKPTPKPVVDEAGSLDNGDFKVTTPDST 240
DB 107 -----VKDNKKKRTKKPTPKPVVDEAGSLDNGDFKVTTPDST 147
QY 241 TQHNKVSFPIKTTAKPINRPSLPNSDTSKETSITVNEKTTVTKSTTTNNKQTSIDG 300
DB 148 TQHNKVSFPIKTTAKPINRPSLPNSDTSKETSITVNEKTTVTKSTTTNNKQTSIDG 207
QY 301 KEKTSABKTOSIEKTSADLAPTSIVLAKPTPKAETTTKGBALTTPKEPTTPKEPAS 360
DB 208 KEKTSABKTOSIEKTSADLAPTSIVLAKPTPKAETTTKGBALTTPKEPTTPKEPAS 267
QY 361 TTPKEPTTPKSAPTTPKEBAPTTPKSAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEB 420
DB 268 TTPKEPTTPKSAPTTPKEBAPTTPKSAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEB 327
QY 421 APTTKSAPTTPKEBAPTTPKKAAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPK 480
DB 328 APTTKSAPTTPKEBAPTTPKKAAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPK 387
QY 481 BPAPTAPKKAAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPT 540
DB 388 BPAPTAPKKAAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPT 447
QY 541 TTKSAAPTTPKESPTTPKEBAPTTPKEBAPTTPKKAAPTTPKEBAPTTPKEBAPTTPKEB 600
DB 448 TTKSAAPTTPKESPTTPKEBAPTTPKEBAPTTPKKAAPTTPKEBAPTTPKEBAPTTPKEB 507
QY 601 APTAKBAPTTPKETAAPTTPKGLTPTTBEKLAAPTTPKEBAPTTPKEBAPTTPKEBAPTTP 660
DB 508 APTAKBAPTTPKETAAPTTPKGLTPTTBEKLAAPTTPKEBAPTTPKEBAPTTPKEBAPTTP 567
QY 661 PEEBAPTTPKAAAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPT 720
DB 568 PEEBAPTTPKAAAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPT 627
QY 721 APTTPKKAAPKELAPTTPKEBPTSTSDKAPTTPKGAAPTTPKEBAPTTPKEBAPTTPK 780
DB 628 APTTPKKAAPKELAPTTPKEBPTSTSDKAPTTPKGAAPTTPKEBAPTTPKEBAPTTPK 687
QY 781 TAPTTLKEBAPTTPKKAAPKELAPTTPKGPSTTSDDKAPTTPKETAAPTTPKEBAPTTPK 840
DB 688 TAPTTLKEBAPTTPKKAAPKELAPTTPKGPSTTSDDKAPTTPKETAAPTTPKEBAPTTPK 747
QY 841 KPAPTTPETPTSEVSTPTTPKEPTTIHNSPESSTBELSAEPTPKALENSPKREGVPT 900
DB 748 KPAPTTPETPTSEVSTPTTPKEPTTIHNSPESSTBELSAEPTPKALENSPKREGVPT 807
QY 901 TKTPATKPEMTTTAKDTERDRTTPETTAAPTTPKATKETAATTEKTESKIRATTTQV 960
DB 808 TKTPATKPEMTTTAKDTERDRTTPETTAAPTTPKATKETAATTEKTESKIRATTTQV 867
QY 961 TSTTTQDTPPKITTLKTTTLAPVYTTTKTITTTTINNKKEEFAKPKDRAATNSKATTPK 1020
DB 868 TSTTTQDTPPKITTLKTTTLAPVYTTTKTITTTTINNKKEEFAKPKDRAATNSKATTPK 927
QY 1021 POKPTKAPKPTSTKPKTPMRVAKPTTPPRKOTSTMPALNTSIAEAMLOTTTRPN 1080
DB 928 POKPTKAPKPTSTKPKTPMRVAKPTTPPRKOTSTMPALNTSIAEAMLOTTTRPN 987
QY 1081 QTPNSKLVENVPKSEDAGAGETPHMLLRPHVMEVETPMDVLPVFNQGIINPMLS 1140
DB 988 QTPNSKLVENVPKSEDAGAGETPHMLLRPHVMEVETPMDVLPVFNQGIINPMLS 1047

QY 1141 DETNINCKGKPYDGLTTLNGLTVARFGHYFWMLSPFSPSPARRITEVMGI PSLPIDVTFT 1200
DB 1048 DETNINCKKPYDGLTTLNGLTVARFGHYFWMLSPFSPSPARRITEVMGI PSLPIDVTFT 1107
QY 1201 RCNCEGKTFPFKDSQYMFRTNDIKDAGYKPIFKGFGGLTQGIYVAALSTAKYKMPESVY 1260
DB 1108 RCNCEGKTFPFKDSQYMFRTNDIKDAGYKPIFKGFGGLTQGIYVAALSTAKYKMPESVY 1167
QY 1261 FPKGGS1QQY1TYKQEPVQKCPGRPALNTPYGGMTQVRRRRFRAIGPGQTHIRIQY 1320
DB 1168 FPKGGS1QQY1TYKQEPVQKCPGRPALNTPYGGMTQVRRRRFRAIGPGQTHIRIQY 1227
QY 1321 SPARLAYODKGYLHNEVKVSIIMRGLPNVVTSAISLPNIRKPDGYDYAFSGDOYVINDV 1380
DB 1228 SPARLAYODKGYLHNEVKVSIIMRGLPNVVTSAISLPNIRKPDGYDYAFSGDOYVINDV 1287
QY 1381 PSRTARAITTRSGQTLISKVWYNCP 1404
DB 1288 PSRTARAITTRSGQTLISKVWYNCP 1311

RESULT 9

US-07-757-022B-50
Sequence 50, Application US/07757022B
Patent No. 6433142

GENERAL INFORMATION:

APPLICANT: Gesner, Thomas G.
APPLICANT: Clark, Stephen C.
APPLICANT: Turner, Katherine
APPLICANT: Hewick, Rodney M.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/757,022B
FILING DATE: 19910910
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Gaert, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 50:
SEQUENCE CHARACTERISTICS:
LENGTH: 1314 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear

MOLECULE TYPE: protein

US-07-757-022B-50

Query Match 92.6%; Score 6971; DB 4; Length 1314;

Best Local Similarity 93.6%; Pred. No. 0;
Matches 1314; Conservative 0; Mismatches 0; Indels 90; Gaps 1;

QY 1 MAMKTLPIYLILLSVFIQOVSSQDSSSCAGRCGEGYSDATCNQYCNHCECPDF 60
DB 1 MAMKTLPIYLILLSVFIQOVSSQDSSSCAGRCGEGYSDATCNQYCNHCECPDF 60
QY 61 KRVCTAELSCGRCFBSFEREGCDACQCKYDKCCPDYESFCAEVHNPTSPSSKAP 120
DB 61 KRVCTA-----HSSVENQSSSSSSSSSSSTIW 66
QY 121 PPSGASQTIKSTTRSKRPKPKTKKVISEBITEHSVENQSSSSSSSSSTIW 180
DB 67 -----HSSVENQSSSSSSSSSSSTIW 90
QY 181 KIKSSKSAANRELQKKLVKDNKKNRTKKKPPVNDVDSGLNDGDKVTPPTST 240
DB 91 KIKSSKSAANRELQKKLVKDNKKNRTKKKPPVNDVDSGLNDGDKVTPPTST 150
QY 241 TQHNKVSIPKITTAKDINRPSLPPNSDSKETSITVNNKTYETKETTITNKQTSIDG 300
DB 151 TQHNKVSIPKITTAKDINRPSLPPNSDSKETSITVNNKTYETKETTITNKQTSIDG 210
QY 301 KEKTSAKETOSIKTSKADLAPSKVLAKPTPAEITTKGPAITTKPEPTTPKBPAS 360
DB 211 KEKTSAKETOSIKTSKADLAPSKVLAKPTPAEITTKGPAITTKPEPTTPKBPAS 270
QY 361 TTPKEPTTKSAPTPPKBPAPTTKSAPTPPKBPAPTTKBPAPTTKBPAPTTKBP 420
DB 271 TTPKEPTTKSAPTPPKBPAPTTKSAPTPPKBPAPTTKBPAPTTKBPAPTTKBP 330
QY 421 APTTKSAPTPPKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTK 480
DB 331 APTTKSAPTPPKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTK 390
QY 481 EPAPTPAKKPAITPKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPT 540
DB 391 EPAPTPAKKPAITPKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPT 450
QY 541 TTKSAPTPPKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBP 600
DB 451 TTKSAPTPPKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBP 510
QY 601 APTAPKBPAPTTPKETAATTPPKLPTTPEKLAATTPKEKAPTTPEELAPTTPEEPTT 660
DB 511 APTAPKBPAPTTPKETAATTPPKLPTTPEKLAATTPKEKAPTTPEELAPTTPEEPTT 570
QY 661 PEEBAPTPPKAAANTKEKAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPT 720
DB 571 PEEBAPTPPKAAANTKEKAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPT 630
QY 721 APTTPKBPAPKELAPTTKKEPTSTSDKPAATTPKGAATTPKBPAPTTKBPAPTTK 780
DB 631 APTTPKBPAPKELAPTTKKEPTSTSDKPAATTPKGAATTPKBPAPTTKBPAPTTK 690
QY 781 TAPTLKBPAPTTPKKAPKELAPTTKGPSTTSDDKAPTTPKETAATTPKBPAPTTPK 840
DB 691 TAPTLKBPAPTTPKKAPKELAPTTKGPSTTSDDKAPTTPKETAATTPKBPAPTTPK 750
QY 841 KPAATTEPTPTTSVSTPTTKKEPTTIHKSPESTPELSAETPALENSPEKGVPT 900
DB 751 KPAATTEPTPTTSVSTPTTKKEPTTIHKSPESTPELSAETPALENSPEKGVPT 810
QY 901 TKTPLATKPEMTTAKKTERDRLTPETTAPKXMTKEATTEKTSKIRATTOV 960
DB 811 TKTPLATKPEMTTAKKTERDRLTPETTAPKXMTKEATTEKTSKIRATTOV 870
QY 961 TSTTODTPEKITTLLKTTLLAPKVTTKTIITTEIMNKPEETAKPRATNSKATTPK 1020

```

Db      871 TSTTTOTTPBKJITLLKTTTLAPKVVTTTKITTTTEIMNKBETAKPKATNSKATTPK 930
Qy      1021 POKPTAPKPKTSTKPKTMRVVRKPKTPTPKMTSTMBELNPTSRIAEMLQTTTRPV 1080
Db      931 POKPTAPKPKTSTKPKTMRVVRKPKTPTPKMTSTMBELNPTSRIAEMLQTTTRPV 990
Qy      1081 QTPNSKLVEVNPKSBDAGAGAGETPHMLLRPHVFMBEVTPDMOYLPRVNOGIIINPMLS 1140
Db      991 QTPNSKLVEVNPKSBDAGAGAGETPHMLLRPHVFMBEVTPDMOYLPRVNOGIIINPMLS 1050
Qy      1141 DETNINCGRKVDLTLRLNGTLVAFRGHYFMMLSPSPSPSPARITBWCISPPIDTVTF 1200
Db      1051 DETNINCGRKVDLTLRLNGTLVAFRGHYFMMLSPSPSPSPARITBWCISPPIDTVTF 1110
Qy      1201 RCNCEGTFPFKQSOYWRFTNDIKDGYPKPIFKGFGGLGQVVAALSTAKYNWPEASY 1260
Db      1111 RCNCEGTFPFKQSOYWRFTNDIKDGYPKPIFKGFGGLGQVVAALSTAKYNWPEASY 1170
Qy      1261 PFKRGSIOQYIKQEBVOKCPGRAPALNYPVYGMTQVRRRFEBAIGPSQTHIRIQY 1320
Db      1171 PFKRGSIOQYIKQEBVOKCPGRAPALNYPVYGMTQVRRRFEBAIGPSQTHIRIQY 1230
Qy      1321 SPARLAVODKGVLANEVKYSILWRGLPNVTSAISLPIRKPDGYDYAFSKQYVNIIDV 1380
Db      1231 SPARLAVODKGVLANEVKYSILWRGLPNVTSAISLPIRKPDGYDYAFSKQYVNIIDV 1290
Qy      1381 PSRTARAITTRSGOTLSKWNVNC 1404
Db      1291 PSRTARAITTRSGOTLSKWNVNC 1314

RESULT 10
US-07-757-022B-142
Sequence 142, Application US/07757022B
Patent No. 6433142
GENERAL INFORMATION:
APPLICANT: Geener, Thomas G.
APPLICANT: Clark, Stephen C.
APPLICANT: Turner, Katherine
APPLICANT: Hewick, Rodney M.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSER: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07757, 022B
FILING DATE: 19910910
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Geer, Luann
REGISTRATION NUMBER: 31,822

```

```

REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 142:
SEQUENCE CHARACTERISTICS:
LENGTH: 1313 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-757-022B-142

Query Match      92.4%; Score 6951.5; DB 4; Length 1313;
Best Local Similarity 93.5%; Pred. No. 0;
Matches 1313; Conservative 0; Mismatches 0; Indels 91; Gaps 2;

Qy      1 MAMKTLPIYLILLISLVFIQVVSODLSACAGCEGYSDATCNCVNCQHYMECCPDF 60
Db      1 MAMKTLPIYLILLISLVFIQVVSQ----- 25
Qy      61 KVCIAELSCRCRCSEFREGRECCDQCKYDKCCPDYSCAEVNPSTSPSSKAP 120
Db      26 -----ELSCRCRCSEFREGRECCDQCKYDKCCPDYSCAE----- 64
Qy      121 PPSGASQRTKSTTKSPKPKKTKVYLSBEITEHVSSENOSSSSSSSSSSSTIM 180
Db      65 -----EHSVSSENOSSSSSSSSSSSSSTIM 89
Qy      181 KIKSSKNSAANELOKLVKONKKNRTKKKTPRPVVDEAGSLDNGDFVYTPDST 240
Db      90 KIKSSKNSAANELOKLVKONKKNRTKKKTPRPVVDEAGSLDNGDFVYTPDST 149
Qy      241 TQHNKVSSTPKITTKAPINRPSLPPNSDTSKETSIVNKEVTKETTTNKQSTNG 300
Db      150 TQHNKVSSTPKITTKAPINRPSLPPNSDTSKETSIVNKEVTKETTTNKQSTNG 209
Qy      301 KEKTSASAKTOSIEKTSANOLAPTSKVLAKPDKAETTKGPAITTPKSPPTTPKSPAS 360
Db      210 KEKTSASAKTOSIEKTSANOLAPTSKVLAKPDKAETTKGPAITTPKSPPTTPKSPAS 269
Qy      361 TTPKEPTPTTIKSAATTPKEPAATTTKSAATTPKEPAATTTKEPAATTTKEP 420
Db      270 TTPKEPTPTTIKSAATTPKEPAATTTKSAATTPKEPAATTTKEPAATTTKEP 329
Qy      421 APTTKSAATTPKEPAATTPKEPAATTPKEPAATTPKEPAATTPKEPAATTPK 480
Db      330 APTTKSAATTPKEPAATTPKEPAATTPKEPAATTPKEPAATTPKEPAATTPK 389
Qy      481 BPAPTAPKPKAPATTPKEPAATTPKEPAATTPKEPAATTPKSAATTPKEPAAT 540
Db      390 BPAPTAPKPKAPATTPKEPAATTPKEPAATTPKEPAATTPKSAATTPKEPAAT 449
Qy      541 TTKSAATTPKEBSPTTIKSAATTPKEPAATTPKEPAATTPKEPAATTPK 600
Db      450 TTKSAATTPKEBSPTTIKSAATTPKEPAATTPKEPAATTPKEPAATTPK 509
Qy      601 APTAKKEPAATTPKEPAATTPKEPAATTPKEPAATTPKEPAATTPKEPAAT 660
Db      510 APTAKKEPAATTPKEPAATTPKEPAATTPKEPAATTPKEPAATTPKEPAAT 569
Qy      661 PEEPAATTPKAAAPTPKEPAATTPKEPAATTPKEPAATTPKEPAATTPK 720
Db      570 PEEPAATTPKAAAPTPKEPAATTPKEPAATTPKEPAATTPKEPAATTPK 629
Qy      721 APTTKKAPKELAPTTTKEPTSTSDKAPATTPKGAATTPKEPAATTPK 780
Db      630 APTTKKAPKELAPTTTKEPTSTSDKAPATTPKGAATTPKEPAATTPK 689
Qy      781 TAPTLKBPAPTPKPAKELAPTTTGPSTSDKAPATTPKGAATTPKEPAATTPK 840
Db      690 TAPTLKBPAPTPKPAKELAPTTTGPSTSDKAPATTPKGAATTPKEPAATTPK 749
Qy      841 KPAPTPEPPTTSEVSTPTTKEPTTIHKSDBESTBELSAEPTPKALENSPKPGVPT 900

```

Db 750 KPAFTPTPTPTTSEVSTPTTKEPTTHHSPTDSTBELSEPTPKALENSPKRGVPT 809
QY 901 TKTPATATPEMTTAKXDKTTERDLTTPBETTAAPRMTKETATTTTEKTESKITATTTQV 960
Db 810 TKTPATATPEMTTAKDKTTERDLTTPBETTAAPRMTKETATTTTEKTESKITATTTQV 869
QY 961 TSTTTODTTPKITTLTKTTLLAPKYTTTKKITTTTEINNKPEPTAKPPDRATNSAATPK 1020
Db 870 TSTTTODTTPKITTLTKTTLLAPKYTTTKKITTTTEINNKPEPTAKPPDRATNSAATPK 929
QY 1021 POKPTAKKPTSTKPKTPMVRKPKTTPPRKMTSTMPBELNPTSLAEMLOTTTSPN 1080
Db 930 POKPTAKKPKPTSTKPKTPMVRKPKTTPPRKMTSTMPBELNPTSLAEMLOTTTSPN 989
QY 1081 QTPNSKLVANPKSDAGABSETPHMLLRPHVMEVETPPMDYLPVYNOGIIINPMLS 1140
Db 990 QTPNSKLVANPKSDAGABSETPHMLLRPHVMEVETPPMDYLPVYNOGIIINPMLS 1049
QY 1141 DETNCKNKPVDGLTTLNGLTVAFRGHYFWMLSPSPSPARRITETVWGI PSPIDVT 1200
Db 1050 DETNCKNKPVDGLTTLNGLTVAFRGHYFWMLSPSPSPARRITETVWGI PSPIDVT 1109
QY 1201 RNCSEKTFPFKDSQYWFPTNDIKDAGYPKPIFKGFGJTGQIYAALSTAKYKMPESVY 1260
Db 1110 RNCSEKTFPFKDSQYWFPTNDIKDAGYPKPIFKGFGJTGQIYAALSTAKYKMPESVY 1169
QY 1261 PFKKGGSIQOYIYKQEPVQKCPRRPALNYPYGGMTQVRRRRFRAIGPSQHTTIRIQY 1320
Db 1170 PFKKGGSIQOYIYKQEPVQKCPRRPALNYPYGGMTQVRRRRFRAIGPSQHTTIRIQY 1229
QY 1321 SPARLAYODKGVLANHEVVSILMRGLPVVYVSATSLPNIRKPDGVDYAFSDOYVNDV 1380
Db 1230 SPARLAYODKGVLANHEVVSILMRGLPVVYVSATSLPNIRKPDGVDYAFSDOYVNDV 1289
QY 1381 PSRTARAITTTRSGOTLSKVMYNCP 1404
Db 1290 PSRTARAITTTRSGOTLSKVMYNCP 1313

RESULT 11
US-07-757-022B-44
Sequence 44, Application US/07757022B
Patent No. 6433142
GENERAL INFORMATION:
APPLICANT: Gesner, Thomas G.
APPLICANT: Clark, Stephen C.
APPLICANT: Turner, Katherine
APPLICANT: Hewick, Rodney M.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/757.022B
FILING DATE: 19910910
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Coert, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 876-1170
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 44:
SEQUENCE CHARACTERISTICS:
LENGTH: 1270 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-757-022B-44

Query Match 89.4%; Score 6727; DB 4; Length 1270;
Best Local Similarity 90.5%; Pred. No. 0;
Matches 1270; Conservative 0; Mismatches 0; Indels 134; Gaps 2;

1 MAMKTLPIYLLILLISVPIQOVSSODLSSCAGRGEGSRATCNQVNHWCPCDF 60
Db 1 MAMKTLPIYLLILLISVPIQOVSSQ----- 25
QY 61 KRVCIAELSCGRGCFESFERGECDCDAQCKYDKCPDYESFCAEVANPTSPSSKAP 120
Db 26 -----ELSCGRGCFESFERGECDCDAQCKYDKCPDYESFCAE----- 65
QY 121 PPSGASQTIKSTYKRSKPKPKKTKKYIESEETIEHSVSENQSSSSSSSSSTIW 180
Db 66 ----- 65
QY 181 KIKSSKNSAANRELQKKLVKONKKNRTKKKPPKPVVDEAGSLDNGDPKVTTPDST 240
Db 66 -----VDNKKKNRTKKKPPKPVVDEAGSLDNGDPKVTTPDST 106
QY 241 TQHNKVSPTSKITTAKEPINSPLPPNSDTSKETSITVKNKETTVEKETTNTKQSTDG 300
Db 107 TQHNKVSPTSKITTAKEPINSPLPPNSDTSKETSITVKNKETTVEKETTNTKQSTDG 166
QY 301 KEKTTSAKETOSIKTSAKDIAPTSKVLAKPTPAETTTGPAITTPKEPTTPPKBPAS 360
Db 167 KEKTTSAKETOSIKTSAKDIAPTSKVLAKPTPAETTTGPAITTPKEPTTPPKBPAS 226
QY 361 TTPKEPTTPPKBPASPTTPPKBPATTTKSAPTTPKEPAPTTPPKBPATTTTKEP 420
Db 227 TTPKEPTTPPKBPASPTTPPKBPATTTKSAPTTPKEPAPTTPPKBPATTTTKEP 286
QY 421 APTTKSAPTTPPKBPATTPPKBPATTPPKBPATTPPKBPATTPPKBPATTPPK 480
Db 287 APTTKSAPTTPPKBPATTPPKBPATTPPKBPATTPPKBPATTPPKBPATTPPK 346
QY 481 EPAPTAKKPAPTTPPKBPATTPPKBPATTPPKBPATTPPKBPATTPPKBPAT 540
Db 347 EPAPTAKKPAPTTPPKBPATTPPKBPATTPPKBPATTPPKBPATTPPKBPAT 406
QY 541 TTKSAPTTPPKBPSTTTTKEPAPTTPPKBPATTPPKBPATTPPKBPATTPPKBP 600
Db 407 TTKSAPTTPPKBPSTTTTKEPAPTTPPKBPATTPPKBPATTPPKBPATTPPKBP 466
QY 601 APTAPKEPAPTTPKETAPTTPPKULTPTTPEKLAAPTTEKDAPTTPPEELAPTTPEEPTPT 660
Db 467 APTAPKEPAPTTPKETAPTTPPKULTPTTPEKLAAPTTEKDAPTTPPEELAPTTPEEPTPT 526
QY 661 PEEBAPTTPKAAADNTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKGTAPTTLKEP 720
Db 527 PEEBAPTTPKAAADNTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKGTAPTTLKEP 586

```

QY 721 APTTPKPAPELAPTTTKEPTSTSDKPAPTPKGIAPTTPKEPAPTTKEPAPTTPKG 780
DB 587 APTTPKPAPELAPTTTKEPTSTSDKPAPTPKGIAPTTPKEPAPTTKEPAPTTPKG 646
QY 781 TAFTTLKEPAPTTPKPAPELAPTTTKEPTSTSDKPAPTTKEPAPTTKEPAPTTKE 840
DB 647 TAFTTLKEPAPTTPKPAPELAPTTTKEPTSTSDKPAPTTKEPAPTTKEPAPTTKE 706
QY 841 KPAPTTPEPTTSTSTPTTKEPTTHKSDESBELASAPTKALENSKEGCVPT 900
DB 707 KPAPTTPEPTTSTSTPTTKEPTTHKSDESBELASAPTKALENSKEGCVPT 766
QY 901 TKTPAATKEMTTAKDKTERDLARTTPTTAAPKATKATTTKEKTESKITATTQV 960
DB 767 TKTPAATKEMTTAKDKTERDLARTTPTTAAPKATKATTTKEKTESKITATTQV 826
QY 961 TSTTTODTTPPKITTLKTTTLAERVTTKTITTTBIMNKESTANPKDANASKATTPK 1020
DB 827 TSTTTODTTPPKITTLKTTTLAERVTTKTITTTBIMNKESTANPKDANASKATTPK 886
QY 1021 POKETAPKPKPTSTTKKPKTMBRVRKKTPTTPRKMTSTWELNPTSRIAEMLQTTTRN 1080
DB 887 POKETAPKPKPTSTTKKPKTMBRVRKKTPTTPRKMTSTWELNPTSRIAEMLQTTTRN 946
QY 1081 QTPNSKLVEVNPKESEDAGAEGETPHMLLRPHVFMPEVTPDMQYLPRVPMQGIINPMLS 1140
DB 947 QTPNSKLVEVNPKESEDAGAEGETPHMLLRPHVFMPEVTPDMQYLPRVPMQGIINPMLS 1006
QY 1141 DEFNINCKRPVDELTLRNGTLVAFRGHFMMLSPSPSPAPARITTEWGISPIDTVP 1200
DB 1007 DEFNINCKRPVDELTLRNGTLVAFRGHFMMLSPSPSPAPARITTEWGISPIDTVP 1066
QY 1201 RNCSEKTPFPKOSQVWRFTNDIKDAGYPRIFKSGGGLTGOVVALSTAKYNWESVY 1260
DB 1067 RNCSEKTPFPKOSQVWRFTNDIKDAGYPRIFKSGGGLTGOVVALSTAKYNWESVY 1126
QY 1261 PFKSGSIOQYIYKQEPVQKCPGRPALNYPVYEMTOVRRRFRERAIQSPOTHTIRQY 1320
DB 1127 PFKSGSIOQYIYKQEPVQKCPGRPALNYPVYEMTOVRRRFRERAIQSPOTHTIRQY 1186
QY 1321 SPARLAYQDKGLHNEVKVSIILWRGLPNVVTSAISLPNIRKPDGYDYAFESKQOYNNIDV 1380
DB 1187 SPARLAYQDKGLHNEVKVSIILWRGLPNVVTSAISLPNIRKPDGYDYAFESKQOYNNIDV 1246
QY 1381 PSTATAITTRSGQTLSTKMYNCP 1404
DB 1247 PSTATAITTRSGQTLSTKMYNCP 1270

```

RESULT 12

```

US-07-757-022B-104
/ Sequence 104, Application US/07757022B
/ Patent No. 6431142
/ GENERAL INFORMATION:
/ APPLICANT: Geener, Thomas G.
/ APPLICANT: Clark, Stephen C.
/ APPLICANT: Turner, Katherine
/ APPLICANT: Hewick, Rodney M.
/ TITLE OF INVENTION: Megakaryocyte Stimulating Factors
/ NUMBER OF SEQUENCES: 143
/ CORRESPONDENCE ADDRESS:
/ ADDRESSER: Genetic Institute, Inc.
/ STREET: 87 Cambridge Park Drive
/ CITY: Cambridge
/ STATE: Massachusetts
/ COUNTRY: U.S.A.
/ ZIP: 02140
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patent Release #1.0, Version #1.25
/ CURRENT APPLICATION DATA:

```

```

/ APPLICATION NUMBER: US/07/757,022B
/ FILING DATE: 19910910
/ CLASSIFICATION: 530
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 07/643,502
/ FILING DATE: 18-JAN-1991
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 07/546,114
/ FILING DATE: 29-JUN-1990
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 07/457,196
/ FILING DATE: 29-DEC-1989
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 07/390,901
/ FILING DATE: 08-AUG-1989
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Geer, Luann
/ REGISTRATION NUMBER: 31,822
/ REFERENCE/DOCKET NUMBER: GI 5190
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (617) 876-1170
/ TELEFAX: (617) 876-5851
/ INFORMATION FOR SEQ ID NO: 104:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 1140 amino acids
/ TYPE: AMINO ACID
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/ US-07-757-022B-104

```

Query Match 80.9%; Score 6085; DB 4; Length 1140;

Best Local Similarity 100.0%; Pred. No. 0; Matches 1140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 MAMKTLPIYLLLLSVFIQVSSQDLSSCAGRCGEYSRDATCNDVNCQHYMECCPDF 60
DB 1 MAMKTLPIYLLLLSVFIQVSSQDLSSCAGRCGEYSRDATCNDVNCQHYMECCPDF 60
QY 61 KAVCTHAEISCKRCESFERGECDDAQCKKYDKCCPDYSEFCABVHNPTSPSSKAP 120
DB 61 KAVCTHAEISCKRCESFERGECDDAQCKKYDKCCPDYSEFCABVHNPTSPSSKAP 120
QY 121 PPSGASQITKSTTKSPRPNNKTKKYIESSEITEHHSVSNQSSSSSSSSSTTW 180
DB 121 PPSGASQITKSTTKSPRPNNKTKKYIESSEITEHHSVSNQSSSSSSSSSTTW 180
QY 181 KIKSSKNSAANRELQKLVKONKKNRTKKPTPKRPVVDKAGSLDNGDFKYTTPDST 240
DB 181 KIKSSKNSAANRELQKLVKONKKNRTKKPTPKRPVVDKAGSLDNGDFKYTTPDST 240
QY 241 TOHNVSTSPKITTAKPINRPSLPPNSDTSKETSILVNKETTVETKETTNNKOTSDG 300
DB 241 TOHNVSTSPKITTAKPINRPSLPPNSDTSKETSILVNKETTVETKETTNNKOTSDG 300
QY 301 KEKTSKAKETOSIKETSKADLAPTSKVLAKPTPKAETTTGKALTTPEKPTTKEPAS 360
DB 301 KEKTSKAKETOSIKETSKADLAPTSKVLAKPTPKAETTTGKALTTPEKPTTKEPAS 360
QY 361 TTPKEPTTITKSAPTTKEPAFTTTSKAPTTTKEPAFTTTSKAPTTTKEPAFTTKEP 420
DB 361 TTPKEPTTITKSAPTTKEPAFTTTSKAPTTTKEPAFTTTSKAPTTTKEPAFTTKEP 420
QY 421 APTTTSKAPTTTKEPAFTTTPKPAFTTTPKEPAFTTTPKEPTTTPKEPAFTTTPK 480
DB 421 APTTTSKAPTTTKEPAFTTTPKPAFTTTPKEPAFTTTPKEPTTTPKEPAFTTTPK 480
QY 481 BPAPTAPKPAFTTTPKEPAFTTTPKEPAFTTTPKEPAFTTTPKEPAFTTTPKEPAFT 540
DB 481 BPAPTAPKPAFTTTPKEPAFTTTPKEPAFTTTPKEPAFTTTPKEPAFTTTPKEPAFT 540
QY 541 TTKSAPTTTKEBSPTTTKEBAPTTPKEBAPTTPKKAFTTTPKEBAPTTPKEBAPTTPK 600
DB 541 TTKSAPTTTKEBSPTTTKEBAPTTPKEBAPTTPKKAFTTTPKEBAPTTPKEBAPTTPK 600

```


Db 630 APTPKKAPKAPKELAPTTKEPTSTSDKPAPTTKGTAPTTKKPAPTTKKPAPTTKG 689
 Qy 781 TAPTLKAPATTPKPKAPKELAPTTKGSTSTSDAPATTKETAATTKKPAPTTK 840
 Db 690 TAPTLKAPATTPKPKAPKELAPTTKGSTSTSDAPATTKETAATTKKPAPTTK 749
 Qy 841 KPAATTPETPTTSEVSTPTTKEPTTIHKSPEDESPELSAETPKALENSPEKGVPT 900
 Db 750 KPAATTPETPTTSEVSTPTTKEPTTIHKSPEDESPELSAETPKALENSPEKGVPT 809
 Qy 901 TKTPAATKPEMTTAAKDKTERDLRTPEETTAAPKMTKETATTEKTESKITATTTQV 960
 Db 810 TKTPAATKPEMTTAAKDKTERDLRTPEETTAAPKMTKETATTEKTESKITATTTQV 869
 Qy 961 TSTTODTTPKKTTLTKTTLLAPKVTTKTITTEIMNKPEETAKKDAATSKATTPK 1020
 Db 870 TSTTODTTPKKTTLTKTTLLAPKVTTKTITTEIMNKPEETAKKDAATSKATTPK 929
 Qy 1021 POKPTAPKPKSTKPKTPMVRKPKTTPTPRKMSTMPELNPTSRIAEAMLTQTTTPN 1080
 Db 930 POKPTAPKPKSTKPKTPMVRKPKTTPTPRKMSTMPELNPTSRIAEAMLTQTTTPN 989
 Qy 1081 QTFNSKLVENVPKSEBAGAEGETPHMLRPVFMPEVTPDMQYLPVPRNOGIIINPMLS 1140
 Db 990 QTFNSKLVENVPKSEBAGAEGETPHMLRPVFMPEVTPDMQYLPVPRNOGIIINPMLS 1049
 RESULT 14
 US-07-757-0228-84
 / Sequence 84, Application US/077570228
 / Patent No. 6433142
 / GENERAL INFORMATION:
 / APPLICANT: Geener, Thomas G.
 / APPLICANT: Clark, Stephen C.
 / APPLICANT: Turner, Katherine
 / APPLICANT: Hewick, Rodney M.
 / TITLE OF INVENTION: Megakaryocyte Stimulating Factors
 / NUMBER OF SEQUENCES: 143
 / CORRESPONDENCE ADDRESS:
 / ADDRESSEE: Genetics Institute, Inc.
 / STREET: 87 Cambridgepark Drive
 / CITY: Cambridge
 / STATE: Massachusetts
 / COUNTRY: U.S.A.
 / ZIP: 02140
 / COMPUTER READABLE FORM:
 / MEDIUM TYPE: Floppy disk
 / COMPUTER: IBM PC compatible
 / OPERATING SYSTEM: PC-DOS/MS-DOS
 / SOFTWARE: Patentin Releasee #1.0, Version #1.25
 / CURRENT APPLICATION DATA:
 / APPLICATION NUMBER: US/07/757,0228
 / FILING DATE: 19910910
 / CLASSIFICATION: 530
 / PRIOR APPLICATION DATA:
 / APPLICATION NUMBER: US 07/643,502
 / FILING DATE: 18-JAN-1991
 / PRIOR APPLICATION DATA:
 / APPLICATION NUMBER: US 07/546,114
 / FILING DATE: 29-JUN-1990
 / PRIOR APPLICATION DATA:
 / APPLICATION NUMBER: US 07/457,196
 / FILING DATE: 29-DEC-1989
 / PRIOR APPLICATION DATA:
 / APPLICATION NUMBER: US 07/390,901
 / FILING DATE: 08-AUG-1989
 / ATTORNEY/AGENT INFORMATION:
 / NAME: Geert, Luann
 / REGISTRATION NUMBER: 31,822
 / REFERENCE/DOCKET NUMBER: GI 5190
 / TELECOMMUNICATION INFORMATION:
 / TELEPHONE: (617) 876-1170
 / TELEFAX: (617) 876-5851

/ INFORMATION FOR SEQ ID NO: 84:
 / SEQUENCE CHARACTERISTICS:
 / LENGTH: 1022 amino acids
 / TYPE: AMINO ACID
 / TOPOLOGY: linear
 / MOLECULE TYPE: protein
 / US-07-757-0228-84
 Query Match 72.1%; Score 5425.5; DB 4; Length 1022;
 Best Local Similarity 91.5%; Pred. No. 0;
 Matches 1020; Conservative 0; Mismatches 2; Indels 93; Gaps 2;
 Db 26 DISSCAGRCGEYSRDATGCDYHMECCPDPRVCYTAELSCGRCFESPEREED 85
 1 DISSCAGRCGEYSRDATGCDYHMECCPDPRVC-ELSCGRCFESPEREED 58
 Qy 86 CDAQCKKYKCCPDYESFCAEYHNPTSPSSKKAAPPSSAQTIKSTGRSPKPKKKT 145
 Db 59 CDAQCKKYKCCPDYESFCAET----- 80
 Qy 146 KAYIESEETHEHSVENOESSSSSSSSSTIWKIKSSKNSANRELQKLVKONK 205
 Db 81 -----AVDNNK 87
 Qy 206 NRTKKKPPKPVVDAGSGLDNGDPKVTTPDSTTHHNVKSTSPKITTAAPINRPSLP 265
 Db 88 NRTKKKPPKPVVDAGSGLDNGDPKVTTPDSTTHHNVKSTSPKITTAAPINRPSLP 147
 Qy 266 PMSDTSKETSLTVNKKETVETKETTNNKQSTDGKEKTSKAKETQSIKTSKADLAPTS 325
 148 PMSDTSKETSLTVNKKETVETKETTNNKQSTDGKEKTSKAKETQSIKTSKADLAPTS 207
 Qy 326 KYLAKPTPAEETTTGPAITTPKEPTTPPKBPASTTPKEPTTIKSAPTTPKEBAPT 385
 Db 208 KYLAKPTPAEETTTGPAITTPKEPTTPPKBPASTTPKEPTTIKSAPTTPKEBAPT 267
 Qy 386 TYSAPTPKEBAPTTPKEBAPTTPKEBAPTTPKSAPTTPKSAPTTPKSAPTTPKBP 445
 Db 268 TYSAPTPKEBAPTTPKEBAPTTPKEBAPTTPKSAPTTPKSAPTTPKSAPTTPKBP 327
 Qy 446 TTPKEBAPTTPKEPTTPKEBAPTTPKEBAPTTPKSAPTTPKSAPTTPKSAPTTPK 505
 Db 328 TTPKEBAPTTPKEPTTPKEBAPTTPKEBAPTTPKSAPTTPKSAPTTPKSAPTTPK 387
 Qy 506 PAPTTPKPSPTTPKEBAPTTPKSAPTTPKSAPTTPKSAPTTPKPSPTTPKEBAPT 565
 Db 388 PAPTTPKPSPTTPKEBAPTTPKSAPTTPKSAPTTPKSAPTTPKPSPTTPKEBAPT 447
 Qy 566 KBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTP 625
 Db 448 KBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTP 507
 Qy 626 PTTPEKLAATTPPEKAPATTPPEELATTPPEEPPTTPEEBAPPTTPKAAANTPEKAPATTP 685
 Db 508 PTTPEKLAATTPPEKAPATTPPEELATTPPEEPPTTPEEBAPPTTPKAAANTPEKAPATTP 567
 Qy 686 KBPATTPKBPATTPKETAAPTPKGINAPTTIKBPAPITTPKBPAPKAPKELAPTTKEPTSTT 745
 Db 568 KBPATTPKBPATTPKETAAPTPKGINAPTTIKBPAPITTPKBPAPKAPKELAPTTKEPTSTT 627
 Qy 746 SDKPAPTPKGAATTPKBPATTPKBPATTPKGAATTPKBPATTPKBPATTPKBPATTP 805
 Db 628 SDKPAPTPKGAATTPKBPATTPKBPATTPKGAATTPKBPATTPKBPATTPKBPATTP 687
 Qy 806 TTKGPTSTSDKAPATTKETAATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTP 865
 Db 688 TTKGPTSTSDKAPATTKETAATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTP 747
 Qy 866 PTTIHKSPDESPELSAETPKALENSPEKGVPTTKTPAATKPEMTTAAKDKTERDLR 925
 Db 748 PTTIHKSPDESPELSAETPKALENSPEKGVPTTKTPAATKPEMTTAAKDKTERDLR 807
 Qy 926 TTPETTAAPKMTKETATTEKTESKITATTTQVTSITTTQDTTPFKITTLTKTLAPKV 985

DB 808 TTPETTAAPKMTKFTATTTETTESKITATTTQVTSSTTQDTTFFKITLTKTTLLAKV 867
QY 966 TTTKTTTITTEIMNKPEETAKPKDAINNSKATTPKQPKTKPKKPTSTKPKTPRVK 1045
DB 868 TTTKTTTITTEIMNKPEETAKPKDAINNSKATTPKQPKTKPKKPTSTKPKTPRVK 927
QY 1046 PTKTTPKMTSTMBELNPTSRIAEAMLOTTTRPQNTNSKLVENPKSEDAAGAGETP 1105
DB 928 PTKTTPKMTSTMBELNPTSRIAEAMLOTTTRPQNTNSKLVENPKSEDAAGAGETP 987
QY 1106 HMLLRPHVFMPEVTEDMDYLPRVNOGIIINPMLS 1140
DB 988 HMLLRPHVFMPEVTEDMDYLPRVNOGIIINPMLS 1022
RESULT 15
US-07-757-022B-74
Sequence 74, Application US/07757022B
Patent No. 6433142
GENERAL INFORMATION:
APPLICANT: Gesner, Thomas G.
APPLICANT: Clark, Stephen C.
APPLICANT: Turner, Katherine
APPLICANT: Hewick, Rodney M.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/757,022B
FILING DATE: 19910910
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Geert, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 876-1170
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 74:
SEQUENCE CHARACTERISTICS:
LENGTH: 1038 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-757-022B-74

Query Match 70.9%; Score 5335; DB 4; Length 1038;
Best Local Similarity 73.9%; Pred. No. 0;
Matches 1038; Conservative 0; Mismatches 0; Indels 366; Gaps 3;

QY 1 MAMKTLPIYLLLSVFIQOVSODLSSGAGRGEGYRDATGCDYNGHMECCPDF 60
DB 1 MAMKTLPIYLLLSVFIQOVSQ----- 25
QY 61 KRVTAEISCKGRGFESEFEREGECDAQCKKYDKCQDYBSFCAEVHNPSPSSKAP 120
DB 26 -----ELSCGRGFESEREGECDAQCKKYDKCQDYBSFCAE----- 65
QY 121 PPSGASQITKSTYRSKPKPKKKTKKYIESEETIEHSVSENQSSSSSSSSSSTIW 180
DB 66 ----- 65
QY 181 KIKSSKSAANRELQKKLVKONKKNRTKKKPKPVVDEAGSGLDNGDFKVTTPDTST 240
DB 66 -----VKONKKNRTKKKPKPVVDEAGSGLDNGDFKVTTPDTST 106
QY 241 TQHNKSTSPKITTAKPINRPSLPPNSDTSKETSLSLVNKEETVETKETTNNKQSTDG 300
DB 107 TQHNKSTSPKITTAKPINRPSLPPNSDTSKETSLSLVNKEETVETKETTNNKQSTDG 166
QY 301 KEKTTSAKETOSIKTSAKDLAPTSKVLAKPTPKAETTTKGPALTTPKEPTTPKBPAS 360
DB 167 KEKTTSAKETOSIKTSAKDLAPTSKVLAKPTPKAETTTKGPALTTPKBPPTTPKBPAS 226
QY 361 TTPKEPTPTTKSAPTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBP 420
DB 227 TTPKEPTPTTKSAPTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBP 286
QY 421 APTTKSAPTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPK 480
DB 287 APTTKSAPTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPK 346
QY 481 EPAPTAAPKAPPTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPT 540
DB 347 EPAPTAAPKAPPTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPT 406
QY 541 TTKAPPTPKESPPTTKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPT 600
DB 407 TTKAPPTPKESPPTTKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPT 466
QY 601 APTAPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPT 660
DB 467 APTAPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPT 526
QY 661 PEBAPPTPKAAANTKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPT 720
DB 527 PEBAPPTPKAAANTKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPT 586
QY 721 APTPKKAPAPKELAPTTTKBPSTTSKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPT 780
DB 587 APTPKKAPAPKELAPTTTKBPSTTSKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPT 646
QY 781 TAPTLKEBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPT 840
DB 647 TAPTLKEBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPT 706
QY 841 KPAPTTTETPTTSEVSTPTTKBPSTTSKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPT 900
DB 707 KPAPTTTETPTTSEVSTPTTKBPSTTSKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPT 766
QY 901 TKTPAATKPEMTTAKKTERDRLTTPETTTAPKMTKETAATTEKTESKITAATTOV 960
DB 767 TKTPAATKPEMTTAKKTERDRLTTPETTTAPKMTKETAATTEKTESKITAATTOV 826
QY 961 TSITTDOTPEKITLTKITTLAPKVTITTKKITTEIMNKPEETAKPKDAINNSKATTPK 1020
DB 827 TSITTDOTPEKITLTKITTLAPKVTITTKKITTEIMNKPEETAKPKDAINNSKATTPK 886
QY 1021 POKPTKAPKPTSTKKEKTMPRVAKPTTTPKMTSTMBELNPTSRIAEAMLOTTTRPN 1080
DB 887 POKPTKAPKPTSTKKEKTMPRVAKPTTTPKMTSTMBELNPTSRIAEAMLOTTTRPN 946

```

Qy 1081 QTENSKLVEVNPKESEDAGAGETPHMLLRPHVFMPEVTPEMDYLPRVPMGIIINPMLS 1140
Db 947 QTENSKLVEVNPKESEDAGAGETPHMLLRPHVFMPEVTPEMDYLPRVPMGIIINPMLS 1006
Qy 1141 DETNIGNGKPVGDLTLRNGTLVAFRGHYFMMLSPFSPSPARRITEWGIPEPIDTVEFT 1200
Db 1007 ----- 1006
Qy 1201 RCNCEGKTFPFKDSQYWRFTNDIKDAGYPRKIFKGFGLTGQIVALSTAKYNWPESVY 1260
Db 1007 ----- 1006
Qy 1261 PFKRGGSIOQYIKQBPVQKCPGRPALNPYVGEMTVQVRRRPERAIGPSQTHIRIQY 1320
Db 1007 ----- 1006
Qy 1321 SPARLAYQDKVLAHEVKVSIILMRGLPNTVTSALSLPNIRKPDGDYDYAFESKDQYNNIDY 1380
Db 1007 -----DOYNNIDY 1014
Qy 1381 PSRTARAITTRSGQTLKVMWYNCP 1404
Db 1015 PSRTARAITTRSGQTLKVMWYNCP 1038

```

Search completed: December 8, 2003, 09:38:09
 Job time : 30 secs

THIS PAGE BLANK (USPTO)

FT /label= Exon_VIII
FT Region /1213..1266
FT /label= Exon_IX
FT Region /1266..1331
FT /label= Exon_X
FT Region /1331..1373
FT /label= Exon_XI
FT Region /1373..1404
FT /label= Exon_XII
PN MO9213075-A1.
PD 06-AUG-1992.
PE 17-JAN-1992; 92MO-US00433.
PF 18-JAN-1991; 91US-0643502.
PR 10-SEP-1991; 91US-0757022.
PS (GENY) GENETICS INST INC.
PX Clark SC, Gesner TG, Hewick RM, Jacobs K, Turner K;
PY MPI; 1992-284660/34.
PZ N-PSDB; AAQ27223.
DR New human mega-karyocyte stimulating factors - for treating
DT Immune deficiencies, cancer, exposure to radiation or drugs,
DT bacterial and viral infections, etc.
XX Claim 1, 2 and 3; Fig 1, 87bp; English.
XX The sequence given is a full length translation from the megakaryocyte
XX stimulating factor (MSF) precursor. The sequence covered by exons II,
XX III and IV encodes megakaryocyte stimulating factor (MSF). This
XX sequence is modified by the addition of an N-terminal sequence encoding
XX a secretory leader, an initiating methionine preceeding exon II and a
XX terminating codon following exon IV. The cDNA sequence given contains
XX sequences derived from human megakaryocyte colony stimulating factor
XX (meg-CSF). Exon I contains the initiating methionine, and encodes a
XX classical mammalian protein secretion signal sequence. The sequence
XX encoding the original meg-CSF includes exons II-IV and is thought to
XX terminate in the region between amino acid residues 134 - 147. The
XX primary transcript of this gene may be cleaved in different ways to
XX yield a family of mRNAs each encoding a different MSF protein. Exons
XX V and VI are thought to be related to the activity of the factor and
XX are also implicated in the stability, folding and processing of the
XX molecule. These exons are also thought to play a role in the observed
XX synergy of MSF with other cytokines. Exons V - XII are believed to be
XX implicated in the processing or folding of the appropriate structure of
XX the resulting factor, ie. one or more of these exons may contain
XX sequences which direct proteolytic cleavage, adhesion, organisation of
XX the cellular matrix or extracellular matrix processing. Both naturally
XX occurring and non-naturally occurring MSFs may be characterised by
XX various combinations of alternatively spliced exons from this sequence,
XX with the exons spliced together in differing orders to form different
XX members of the MSF family.
XX (Updated on 25-MAR-2003 to correct PN field.)
SQ Sequence 1404 AA;
Query Match 100.0%; Score 7526; DB 13; Length 1404;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1404; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 121 PPSGASOTIKSTTRSPKPKKKTKKYIESEETIEHSVSENOSSSSSSSSSTTW 180
DB 121 PPSGASOTIKSTTRSPKPKKKTKKYIESEETIEHSVSENOSSSSSSSSSTTW 180
QY 181 KIKSSKNSAANRELQKLVKNDKNNKTRKKKTPKPVVDEAGSLDNGDKVTPDTST 240
DB 181 KIKSSKNSAANRELQKLVKNDKNNKTRKKKTPKPVVDEAGSLDNGDKVTPDTST 240
QY 241 TQHNKSTSPKLTAKKPINRPSLPNSDTSKETSILVNKETTVEETTTNNQSTDG 300
DB 241 TQHNKSTSPKLTAKKPINRPSLPNSDTSKETSILVNKETTVEETTTNNQSTDG 300
QY 301 KEKTTSAKETOSIKETSADLAPTSKYLAKPTPAETTTKGPALTTPKEPPTTPKEBAS 360
DB 301 KEKTTSAKETOSIKETSADLAPTSKYLAKPTPAETTTKGPALTTPKEPPTTPKEBAS 360
QY 361 TTPKEPPTTTIKSAPTTPKKEBAPTTTTSKAPTTPKKEBAPTTTKEBAPTTTKEP 420
DB 361 TTPKEPPTTTIKSAPTTPKKEBAPTTTTSKAPTTPKKEBAPTTTKEBAPTTTKEP 420
QY 421 APTTTKAPTTPKKEBAPTTPKKAPTTTPKEBAPTTTPKEBAPTTTPKEBAPTTTPK 480
DB 421 APTTTKAPTTPKKEBAPTTPKKAPTTTPKEBAPTTTPKEBAPTTTPKEBAPTTTPK 480
QY 481 EPAPTAAPKAPPTPKSAPPTTPKEBAPTTTKEBSPPTTPKEBAPTTTTSKAPTTTKEBAPT 540
DB 481 EPAPTAAPKAPPTPKSAPPTTPKEBAPTTTKEBSPPTTPKEBAPTTTTSKAPTTTKEBAPT 540
QY 541 TTKSAPPTPKESPPTTKEBAPTTTPKEBAPTTTPKEBAPTTTPKEBAPTTTPKEBAPT 600
DB 541 TTKSAPPTPKESPPTTKEBAPTTTPKEBAPTTTPKEBAPTTTPKEBAPTTTPKEBAPT 600
QY 601 APTPKAPAPPTPKETAAPTTPKKLTPTTPELALTTPEKAPTTPELAPTTPEEPPTT 660
DB 601 APTPKAPAPPTPKETAAPTTPKKLTPTTPELALTTPEKAPTTPELAPTTPEEPPTT 660
QY 661 PEBAPPTPKAAANTKEBAPTTTPKEBAPTTTPKEBAPTTTPKEBAPTTTPKGTATTTKEP 720
DB 661 PEBAPPTPKAAANTKEBAPTTTPKEBAPTTTPKEBAPTTTPKEBAPTTTPKGTATTTKEP 720
QY 721 APTTPKKAPKELAPTTTKEBPTSTSDKPAPTTPKGTAPTTPKKAPPTTPKEBAPT 780
DB 721 APTTPKKAPKELAPTTTKEBPTSTSDKPAPTTPKGTAPTTPKKAPPTTPKEBAPT 780
QY 781 TAPTTLKEBAPTTTPKKAAPKELAPTTTSGPTSTSDKPAPTTPKETAAPTTPKKEBAPT 840
DB 781 TAPTTLKEBAPTTTPKKAAPKELAPTTTSGPTSTSDKPAPTTPKETAAPTTPKKEBAPT 840
QY 841 KPAPTTEPPTTSEVSTPTTKEPPTTIHKSPESTPELSAETPALENSPEKGVPT 900
DB 841 KPAPTTEPPTTSEVSTPTTKEPPTTIHKSPESTPELSAETPALENSPEKGVPT 900
QY 901 TKTPAATKPEMTTAKODTTERDLRTTPETTTAAAPKMTKETATTTTEKTTESKITATTTTQV 960
DB 901 TKTPAATKPEMTTAKODTTERDLRTTPETTTAAAPKMTKETATTTTEKTTESKITATTTTQV 960
QY 961 TSTTQOTTPPKITLTKTTLAPKVTYTTTKTITTEIMNKPEEPAKODATNSKATTPK 1020
DB 961 TSTTQOTTPPKITLTKTTLAPKVTYTTTKTITTEIMNKPEEPAKODATNSKATTPK 1020
QY 961 TSTTQOTTPPKITLTKTTLAPKVTYTTTKTITTEIMNKPEEPAKODATNSKATTPK 1020
DB 961 TSTTQOTTPPKITLTKTTLAPKVTYTTTKTITTEIMNKPEEPAKODATNSKATTPK 1020
QY 1021 POKPTKAPKPTSTSKKPTKPRVAKPTTPTPRKMTSTMBELNFTSIAEMLQTTTRPN 1080
DB 1021 POKPTKAPKPTSTSKKPTKPRVAKPTTPTPRKMTSTMBELNFTSIAEMLQTTTRPN 1080
QY 1081 QTPNSKLVEVNPKSEDAGAGETPHMLLRPHVMPVETPDMDYLPRVNPQGIINPMLS 1140
DB 1081 QTPNSKLVEVNPKSEDAGAGETPHMLLRPHVMPVETPDMDYLPRVNPQGIINPMLS 1140
QY 1141 DETNINCGKPYDGLTTLRNGTLVAFRGHYFMWLSPEPSPAPARITTEWGISPIDVFT 1200
DB 1141 DETNINCGKPYDGLTTLRNGTLVAFRGHYFMWLSPEPSPAPARITTEWGISPIDVFT 1200
QY 1201 RCNCEGKTFPPKDSQVYRFTNDIKDAGYPRIFKFGGLNGQIVAAALSTAKYKWPESVY 1260

DB 1201 RCNCEGTEFFPKDSQYWRFTNDIKDAGYKPIFKGFGGLTGQVVALSTAKKNMPESEY 1260

QY 1261 PFRFGSIOQYTYKOPVOKCGRPALVYVGYEMTOVRRRRFEALIPSGHTIRIOY 1320

DB 1261 PFRFGSIOQYTYKOPVOKCGRPALVYVGYEMTOVRRRRFEALIPSGHTIRIOY 1320

QY 1321 SPARLAYQDKVLAHNEVKVSLMRGLPNVYTSALPNIRKPDGYDYAFSKDQYNNIDV 1380

DB 1321 SPARLAYQDKVLAHNEVKVSLMRGLPNVYTSALPNIRKPDGYDYAFSKDQYNNIDV 1380

QY 1381 PSRTARAITTRSGQTLISKVYNCP 1404

DB 1381 PSRTARAITTRSGQTLISKVYNCP 1404

RESULT 2

AAB60568

ID AAB60568 standard; Protein; 1404 AA.

AC AAB60568;

XX 27-APR-2001 (first entry)

DT Human megakaryocyte stimulating factor (MSF, CACP).

DE Human megakaryocyte stimulating factor (MSF, CACP).

XX Human; CACP protein; campodactyly-arthropathy-coxa vara-pericarditis;

KM MSF; megakaryocyte stimulating factor; synovial lubricant;

KM chromosome 1q25-31; osteoarthritis; joint lubrication; osteopathic;

KW antiarthritis.

OS Homo sapiens.

XX MO200107068-A1.

PN 01-FEB-2001.

PD 21-JUL-2000; 2000MO-US20002.

PF 23-JUL-1999; 99US-0145328.

PR 19-JUL-2000; 2000US-0145328.

XX (UYCA-) UNIV CASE WESTERN RESERVE.

PA Warman ML;

PI WPI; 2001-182721/18.

DR New composition comprising the campodactyly-arthropathy-coxa

PT vara-pericarditis protein in combination with an anesthetic, useful for

PT treating osteoarthritis, or as lubricants of tissue and joints.

XX Example 1; Page -; 34p; English.

XX The invention relates to a method of treating osteoarthritis via the

CC administration of a composition comprising the campodactyly-arthropathy-

CC coxa vara-pericarditis (CACP) protein, or portions of the CACP protein.

CC The composition may further comprise a local anesthetic. The composition

CC of the invention may be administered via intra-articular or intravenous

CC injection. The human CACP protein is identified in the invention as

CC being megakaryocyte stimulating factor (MSF). The gene encoding

CC CACP protein (MSF) is located on chromosome 1q25-31, and mutations in

CC this gene are responsible for the heritable disorder campodactyly-

CC arthropathy-coxa vara-pericarditis, in which patients have synovial

CC hyperplasia without evidence of inflammation. CACP protein (MSF)

CC acts as a synovial lubricant, and can be used to lubricate tissue and

CC joints in the treatment of osteoarthritis. The composition may be

CC applied to reduce the symptoms of osteoarthritis (e.g., joint pain,

CC loss of range of movement or joint damage). The present sequence

CC represents human megakaryocyte stimulating factor (MSF, CACP protein).

CC Note: This sequence is not given in its entirety in figure 4 of the

CC specification, although a Genbank accession number was given. This

CC sequence was therefore obtained from Genbank (U70316).

XX SQ Sequence 1404 AA;

Query Match 100.0%; Score 7526; DB 22; Length 1404;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1404; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAMKLPYLLLLSVFIQOVSSODLSSCAGRCGEYSRDTCCDNCQHYMECCPDF 60

DB 1 MAMKLPYLLLLSVFIQOVSSODLSSCAGRCGEYSRDTCCDNCQHYMECCPDF 60

QY 61 KRVCTAELSCGRCESEFERGECDDAQCKKYDCCPDYESFCAEVNPTSPSSKAP 120

DB 61 KRVCTAELSCGRCESEFERGECDDAQCKKYDCCPDYESFCAEVNPTSPSSKAP 120

QY 121 PPSGASQITIKSTTKKSPKPPNKKTKKVISEBEITEHHSVENQSSSSSSSSSTIW 180

DB 121 PPSGASQITIKSTTKKSPKPPNKKTKKVISEBEITEHHSVENQSSSSSSSSSTIW 180

QY 181 KIKSSKNSAANRELQKLVKNDKKNRTKKKPTPKPVVDEAGSLDNGDFVYTPDST 240

DB 181 KIKSSKNSAANRELQKLVKNDKKNRTKKKPTPKPVVDEAGSLDNGDFVYTPDST 240

QY 241 TQHNKYSTSPKITTAKPINRPSLPPNSDTSKETSILVANKETVETKETTNNKQSTDG 300

DB 241 TQHNKYSTSPKITTAKPINRPSLPPNSDTSKETSILVANKETVETKETTNNKQSTDG 300

QY 301 KEKTTSAKETOSIEKTSADLAFTSKVLAKPTPKAETTTKGPALTTPKEPTTPKEPAS 360

DB 301 KEKTTSAKETOSIEKTSADLAFTSKVLAKPTPKAETTTKGPALTTPKEPTTPKEPAS 360

QY 361 TTPKEPTTTPKSAPTTKEBPATTTKSAPTTKEBPATTTKEBPATTTKEBPATTTKEP 420

DB 361 TTPKEPTTTPKSAPTTKEBPATTTKSAPTTKEBPATTTKEBPATTTKEBPATTTKEP 420

QY 421 APTTTSKAPTTKEBPATTTKPKKAPTTKEBPATTTKEBPATTTKEBPATTTKEBPAT 480

DB 421 APTTTSKAPTTKEBPATTTKPKKAPTTKEBPATTTKEBPATTTKEBPATTTKEBPAT 480

QY 481 EPAPAPKAPKAPTTKEBPATTTKEBPATTTKEBPATTTKEBPATTTKEBPATTTKEBP 540

DB 481 EPAPAPKAPKAPTTKEBPATTTKEBPATTTKEBPATTTKEBPATTTKEBPATTTKEBP 540

QY 541 TTKSAPTTKESPSPTTKEBPATTTKEBPATTTKEBPATTTKEBPATTTKEBPATTTKEP 600

DB 541 TTKSAPTTKESPSPTTKEBPATTTKEBPATTTKEBPATTTKEBPATTTKEBPATTTKEP 600

QY 601 APTAPKEBPATTTKEBPATTTKEBPATTTKEBPATTTKEBPATTTKEBPATTTKEP 660

DB 601 APTAPKEBPATTTKEBPATTTKEBPATTTKEBPATTTKEBPATTTKEBPATTTKEP 660

QY 661 PEEBPATTPKAAAPNTKEBPATTTKEBPATTTKEBPATTTKEBPATTTKEBPATTTKEP 720

DB 661 PEEBPATTPKAAAPNTKEBPATTTKEBPATTTKEBPATTTKEBPATTTKEBPATTTKEP 720

QY 721 APTPKKAPKAPKAPTTKEBPATTTKEBPATTTKEBPATTTKEBPATTTKEBPATTTKEP 780

DB 721 APTPKKAPKAPKAPTTKEBPATTTKEBPATTTKEBPATTTKEBPATTTKEBPATTTKEP 780

QY 781 TAPTLKBPATTTKEBPATTTKEBPATTTKEBPATTTKEBPATTTKEBPATTTKEBPAT 840

DB 781 TAPTLKBPATTTKEBPATTTKEBPATTTKEBPATTTKEBPATTTKEBPATTTKEBPAT 840

QY 841 KPAPTPPTPTPTSEVSTPTTKKPTTHKSPESTBELSABPTPKALNSPKKGPVPT 900

DB 841 KPAPTPPTPTPTSEVSTPTTKKPTTHKSPESTBELSABPTPKALNSPKKGPVPT 900

QY 901 TKTBAATKEPTTAKKOTTERDRTTPTTAAPKMKETATTTKETSKITATTTQV 960

DB 901 TKTBAATKEPTTAKKOTTERDRTTPTTAAPKMKETATTTKETSKITATTTQV 960

QY 961 TSTTQODTTPPKITTLKTTTLAPKVTTKKTTITTEINNKKESTAKPDORATNSKATTPK 1020

DB 961 TSTTDDTPTKLTITLTKTTTLAFAKVTITTKITITTEIMNKKPEETAKPRORATNSKATTPK 1020
QY 1021 POKPTKAPKKTSTSTKKTMTKPRVRKPKTTPPRKKTSTIMPELNPISRIAEAMLQTTTTPN 1080
DB 1021 POKPTKAPKKTSTSTKKTMTKPRVRKPKTTPPRKKTSTIMPELNPISRIAEAMLQTTTTPN 1080
QY 1081 QTPNSKLVENVPKSDAGAEGETHMLLRPHVFEVETPPMDVLPVYNOGIIINPMLS 1140
DB 1081 QTPNSKLVENVPKSDAGAEGETHMLLRPHVFEVETPPMDVLPVYNOGIIINPMLS 1140
QY 1141 DETNINCKPVDGLTTLNGLTVAFRGHYFWMLSPFSPSPPARITTEVWGIPSPIDVET 1200
DB 1141 DETNINCKPVDGLTTLNGLTVAFRGHYFWMLSPFSPSPPARITTEVWGIPSPIDVET 1200
QY 1201 RCONEGKTFEFDQSYMRFTNDIKDAGYPKPIFKFGSLTQIYVAALSTAKYKMPESVY 1260
DB 1201 RCONEGKTFEFDQSYMRFTNDIKDAGYPKPIFKFGSLTQIYVAALSTAKYKMPESVY 1260
QY 1261 FFKGSGSIQYIYKQEPVQKCPGRPALNYPVYGMTQVRRRRFRAIGPSQTHTRIQY 1320
DB 1261 FFKGSGSIQYIYKQEPVQKCPGRPALNYPVYGMTQVRRRRFRAIGPSQTHTRIQY 1320
QY 1321 SPARLAYODKGVLANEVKVSILMRGLPNVVTSAISLPIRKPDGYDYAFSKDQYINIDV 1380
DB 1321 SPARLAYODKGVLANEVKVSILMRGLPNVVTSAISLPIRKPDGYDYAFSKDQYINIDV 1380
QY 1381 PSRTARATITRSQGLTSKVMYNCP 1404
DB 1381 PSRTARATITRSQGLTSKVMYNCP 1404

RESULT 3

AAB29773
ID AAB29773 standard; Procein; 1404 AA.

AC AAB29773;

DT 28-FEB-2001 (first entry)

DE Human megakaryocyte stimulating factor (MSF), SEQ ID NO.1.

XX Human MSF; megakaryocyte stimulating factor; tribonecin;

KW alternative splicing; joint boundary lubricant; O-linked oligosaccharide;

KM osteoarthritis; tribosupplementation; tissue adhesion inhibition;

KM friction coefficient reduction; gene therapy; antiarthritic;

XX osteopathic.

XX Homo sapiens.

XX OS

XX MO200064930-A2.

XX PD 02-NOV-2000.

XX PF 24-APR-2000; 2000MO-US10953.

XX PR 23-APR-1999; 99US-0298970.

XX XX (RHOD-) RHODE ISLAND HOSPITAL LIFESPAN PARTNER.

XX PI Jay GD;

XX DR MPI; 2001-024673/03.

XX DR N-PSDB; AAC81498.

XX PT Novel tribonecin polypeptide useful as lubricant for treating

XX PS osteoarthritis, comprises O-linked lubricating moiety

XX Claim 3; Page 7; 47pp; English.

CC repeats of a motif having at least 50% identity to the sequence KSPAPPT
CC (AAB29773). The invention also relates to a nucleic acid encoding a
CC human MSF-derived tribonecin; a biocompatible composition comprising a
CC human tribonecin for inhibiting tissue adhesion formation; and a method
CC of diagnosing osteoarthritis or a predisposition to osteoarthritis by
CC measuring the amount of MSF or its fragment in a biological sample of a
CC mammal, wherein an increased amount of MSF compared to a control
CC indicates the presence of or predisposition to developing
CC osteoarthritis. The tribonecin and DNA encoding it are useful in the
CC treatment of osteoarthritis, where they may be used for lubricating
CC mammalian joints, such as articulating joints of humans, dogs or horses.
CC The tribonecin, when formulated as a membrane, foam, gel or fibre, is
CC useful for inhibiting adhesion between two surfaces such as the injured
CC tissues of a mammal, where the injury is caused by a surgical insertion
CC or trauma, or an artificial device e.g., an orthopaedic implant. In
CC particular, one of the surfaces is pericardial tissue. DNA encoding a
CC tribonecin may be used in gene therapy. The present sequence represents
CC human MSF.

XX SQ Sequence 1404 AA;

Query Match 100.0%; Score 7526; DB 22; Length 1404;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1404; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAMKTLPIYLILLSVFIQVSSQDLSGCGRGEGYSDATGNCNCHWECPCDF 60

DB 1 MAMKTLPIYLILLSVFIQVSSQDLSGCGRGEGYSDATGNCNCHWECPCDF 60

QY 61 KRVCIAELSCGRCFESFERGECDCDQCKYDKCCPDYESFCAEVHNPTSPSSKAP 120

DB 61 KRVCIAELSCGRCFESFERGECDCDQCKYDKCCPDYESFCAEVHNPTSPSSKAP 120

QY 121 PPSGASQTIKSTYKRSKPKPKKTKYVISEEITTEHSVSENDESSSSSSSSSTIW 180

DB 121 PPSGASQTIKSTYKRSKPKPKKTKYVISEEITTEHSVSENDESSSSSSSSSTIW 180

QY 181 KIKSSKNSANRELQKLVKNDKNTKKKPPKPVVDEAGSLGNDGPKVTPPTST 240

DB 181 KIKSSKNSANRELQKLVKNDKNTKKKPPKPVVDEAGSLGNDGPKVTPPTST 240

QY 241 TOHNVSTSPKITAKINPAPSLPNSDTSKETSILVNNKETTETTTNKTSTDG 300

DB 241 TOHNVSTSPKITAKINPAPSLPNSDTSKETSILVNNKETTETTTNKTSTDG 300

QY 301 KEKTSAKETOSIKTSKADLAPTSKVLAKPTPAETTTGPAITTPKEPTTPKBPAS 360

DB 301 KEKTSAKETOSIKTSKADLAPTSKVLAKPTPAETTTGPAITTPKEPTTPKBPAS 360

QY 361 TTPKEPTTPKBPAPPTTPKBPAPPTTPKBPAPPTTPKBPAPPTTPKBPAPTTTKEP 420

DB 361 TTPKEPTTPKBPAPPTTPKBPAPPTTPKBPAPPTTPKBPAPPTTPKBPAPTTTKEP 420

QY 421 APITTKSAPPTPKBPAPPTTPKBPAPPTTPKBPAPPTTPKBPAPPTTPKBPAPTTTKEP 480

DB 421 APITTKSAPPTPKBPAPPTTPKBPAPPTTPKBPAPPTTPKBPAPPTTPKBPAPTTTKEP 480

QY 481 EPAPTAPEKBPAPPTTPKBPAPPTTPKBPAPPTTPKBPAPPTTPKBPAPTTTKEP 540

DB 481 EPAPTAPEKBPAPPTTPKBPAPPTTPKBPAPPTTPKBPAPPTTPKBPAPTTTKEP 540

QY 541 TTKSAPPTPKBPAPPTTPKBPAPPTTPKBPAPPTTPKBPAPPTTPKBPAPTTTKEP 600

DB 541 TTKSAPPTPKBPAPPTTPKBPAPPTTPKBPAPPTTPKBPAPPTTPKBPAPTTTKEP 600

QY 601 APAPKBPAPPTPKBPAPPTTPKBPAPPTTPKBPAPPTTPKBPAPPTTPKBPAPTTTKEP 660

DB 601 APAPKBPAPPTPKBPAPPTTPKBPAPPTTPKBPAPPTTPKBPAPPTTPKBPAPTTTKEP 660

QY 661 PEEPAPPTPKBPAPPTTPKBPAPPTTPKBPAPPTTPKBPAPPTTPKBPAPTTTKEP 720

DB 661 PEEPAPPTPKBPAPPTTPKBPAPPTTPKBPAPPTTPKBPAPPTTPKBPAPTTTKEP 720

```

QY 721 APTTPKKPAKELAPTTTKEPTSTSDKPAPTTPKEAPATTTPKEAPATTPKG 780
DB 721 APTTPKKPAKELAPTTTKEPTSTSDKPAPTTPKEAPATTTPKEAPATTPKG 780
QY 781 TAPTTTKEAPATTTPKEAPELAPTTTKEPTSTSDKPAPTTPKEAPATTTPKEAPATTPK 840
DB 781 TAPTTTKEAPATTTPKEAPELAPTTTKEPTSTSDKPAPTTPKEAPATTTPKEAPATTPK 840
QY 841 KPAPTTPEPTPTTSEVSTPTTKEPTTHKSPDESPBELSABPTPALNSKEGVP 900
DB 841 KPAPTTPEPTPTTSEVSTPTTKEPTTHKSPDESPBELSABPTPALNSKEGVP 900
QY 901 TKTPTAATKPEMTTAAOKTERDLRTTPTTAAAPKATKEATTTTEKTESKITATTTTQV 960
DB 901 TKTPTAATKPEMTTAAOKTERDLRTTPTTAAAPKATKEATTTTEKTESKITATTTTQV 960
QY 961 TSTTTTODTTPPKITTLTKTTTLAPKVTTTKTITTEIMNKPEETAKPKDRAINSKATTPK 1020
DB 961 TSTTTTODTTPPKITTLTKTTTLAPKVTTTKTITTEIMNKPEETAKPKDRAINSKATTPK 1020
QY 1021 POKPTAPKKPTSTTKPKTPMVRKPKTTPTPRKMTSTMBELNPTSRIAEAMLTQTTTRPV 1080
DB 1021 POKPTAPKKPTSTTKPKTPMVRKPKTTPTPRKMTSTMBELNPTSRIAEAMLTQTTTRPV 1080
QY 1081 QTENSKLVEVNPKESEDAGAGEETPHMLLRPHYFMEVEVTPDMOYLPRVPNOGIIINPMS 1140
DB 1081 QTENSKLVEVNPKESEDAGAGEETPHMLLRPHYFMEVEVTPDMOYLPRVPNOGIIINPMS 1140
QY 1141 DEFNICKGKPVDTLTIRNGTLVAFRGHYFMWLSPPSPSPARRITEWGIPIPIDTVFT 1200
DB 1141 DEFNICKGKPVDTLTIRNGTLVAFRGHYFMWLSPPSPSPARRITEWGIPIPIDTVFT 1200
QY 1201 RCNCEKTPFPKSOVMRFTNDIKDAGYPRKIFKRGGLTGOVVAALSTAKYNMESVY 1260
DB 1201 RCNCEKTPFPKSOVMRFTNDIKDAGYPRKIFKRGGLTGOVVAALSTAKYNMESVY 1260
QY 1261 FPKRGSSIQOYIYKQEBVQKCPGRPALNPYVGEVTVRRRFRFBAIGPSQTHTRIOY 1320
DB 1261 FPKRGSSIQOYIYKQEBVQKCPGRPALNPYVGEVTVRRRFRFBAIGPSQTHTRIOY 1320
QY 1321 SPARLAYODKGVLANEVKVSILMRGLPNVVTSAISLPNIRKPDGYDYAFSKOQYNNIDV 1380
DB 1321 SPARLAYODKGVLANEVKVSILMRGLPNVVTSAISLPNIRKPDGYDYAFSKOQYNNIDV 1380
QY 1381 PSRTAARITTRSGOTLSKMYNCP 1404
DB 1381 PSRTAARITTRSGOTLSKMYNCP 1404

```

```

RESULT 4
AAU32262
ID AAU32262 standard; Protein; 1415 AA.
XX
AC AAU32262;
XX
DT 18-DEC-2001 (first entry)
XX
DE Novel human secreted protein #2753.
XX
KW Human; vaccination; gene therapy; nutritional supplement;
KW stem cell proliferation; haematopoiesis; nerve tissue regeneration;
KW immune suppression; immune stimulation; anti-inflammatory; leukaemia.
XX
OS Homo sapiens.
XX
PN WO200179449-A2.
XX
PD 25-OCT-2001.
XX
PP 16-APR-2001; 2001MO-US08656.
XX
PR 18-APR-2000; 2000US-0552929.
PR 26-JAN-2001; 2001US-0770160.

```

```

XX (HYSE-) HYSBQ INC.
PA
PI Tang YT, Liu C, Dmanac RT;
XX MPI; 2001-611725/70.
XX
PT Nucleic acids encoding a range of human polypeptides, useful in genetic
PT vaccination, testing and therapy -
XX
PS Claim 20; Page 573; 765pp; English.
XX
CC The invention relates to novel human secreted polypeptides. The
CC polypeptides and antibodies to the polypeptides are useful for
CC determining the presence of or predisposition to a disease associated
CC with altered levels of polypeptide. The polypeptides are also useful for
CC identifying agents (agonists and antagonists) that bind to them. Cells
CC expressing the proteins are useful for identifying a therapeutic agent
CC for use in treatment of a pathology related to aberrant expression or
CC physiological interactions of the polypeptide. Vectors comprising
CC the nucleic acids encoding the polypeptides and cells genetically
CC engineered to express them are also useful for producing the proteins.
CC The proteins are useful in genetic vaccination, testing and
CC therapy, and can be used as nutritional supplements. They may be used to
CC increase stem cell proliferation; to regulate haematopoiesis; and in
CC bone, cartilage, tendon and/or nerve tissue growth or regeneration;
CC immune suppression and/or stimulation; as anti-inflammatory agents; and
CC in treatment of leukaemias. AAU29510-AAU33304 represent the amino acid
CC sequences of novel human secreted proteins of the invention.
XX
SQ Sequence 1415 AA;

```

```

Query Match 100.0%; Score 7523; DB 22; Length 1415;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1403; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAMKTLPIYLALLLVFVIVQVSSODLSSCAGRCGEYSRDATCNDYVCOHYMECCP 60
DB 12 MAMKTLPIYLALLLVFVIVQVSSODLSSCAGRCGEYSRDATCNDYVCOHYMECCP 71

QY 61 KEVTAEELSCRCRCESEFERGECDDAQCKKYDKCDEYSEFCAEVNHTSPSSKAP 120
DB 72 KAVCTAEELSCRCRCESEFERGECDDAQCKKYDKCDEYSEFCAEVNHTSPSSKAP 131

QY 121 PPSGASQITTKSTTKSPRPNNKTKVIESREITEHSVSNQSSSSSSSSSTTW 180
DB 132 PPSGASQITTKSTTKSPRPNNKTKVIESREITEHSVSNQSSSSSSSSSTTW 191

QY 181 KIKSSKNSAANRELQKDKVNDKONKQRTKKPTPKPVVDAAGSLDNGDFVYTPPDST 240
DB 192 KIKSSKNSAANRELQKDKVNDKONKQRTKKPTPKPVVDAAGSLDNGDFVYTPPDST 251

QY 241 TQHNKYSTSPKITTAAPINRPSLPNSDTSKETSILTANKETVETKETTITNKQSTDG 300
DB 252 TQHNKYSTSPKITTAAPINRPSLPNSDTSKETSILTANKETVETKETTITNKQSTDG 311

QY 301 KEKTTSAKETOSIEKTSAXDLAPISKVLAKPTPKAETTTKGPALTTPKEPTTPKEPAS 360
DB 312 KEKTTSAKETOSIEKTSAXDLAPISKVLAKPTPKAETTTKGPALTTPKEPTTPKEPAS 371

QY 361 TTPKEPTPTTKSAPTTKEBAPTTPKSAPTTPKBPATTPKBPATTPKBPATTPKBP 420
DB 372 TTPKEPTPTTKSAPTTKEBAPTTPKSAPTTPKBPATTPKBPATTPKBPATTPKBP 431

QY 421 APTTTSAPTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBP 480
DB 432 APTTTSAPTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBP 491

QY 481 BPAPTAAPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPAT 540
DB 492 BPAPTAAPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPAT 551

QY 541 TTKSAPTTPKBPSTTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBP 600

```

```
Db 552 TTKAPPTPKESPPTTKEBPATTPKEBPATTPKEBPATTPKEBPATTPKEBPATTPK 611
Qy 601 APTAPKEBPATTPKEBPATTPKEBPATTPKEBPATTPKEBPATTPKEBPATTPK 660
Db 612 APTAPKEBPATTPKEBPATTPKEBPATTPKEBPATTPKEBPATTPKEBPATTPK 671
Qy 661 PEEBPATTPKAAAPTPKEBPATTPKEBPATTPKEBPATTPKEBPATTPKEBPATTPK 720
Db 672 PEEBPATTPKAAAPTPKEBPATTPKEBPATTPKEBPATTPKEBPATTPKEBPATTPK 731
Qy 721 APTTPKEBPAPKEBPATTPKEBPATTPKEBPATTPKEBPATTPKEBPATTPKEBP 780
Db 732 APTTPKEBPAPKEBPATTPKEBPATTPKEBPATTPKEBPATTPKEBPATTPKEBP 791
Qy 781 TAPPTLKEBPATTPKEBPATTPKEBPATTPKEBPATTPKEBPATTPKEBPATTPK 840
Db 792 TAPPTLKEBPATTPKEBPATTPKEBPATTPKEBPATTPKEBPATTPKEBPATTPK 851
Qy 841 KPAPTTPPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 900
Db 852 KPAPTTPPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 911
Qy 901 TKTPTAPATPEMTTAKDKTTERDLRTPEBTAAAPKMTKETATTEKTESKITATTTQV 960
Db 912 TKTPTAPATPEMTTAKDKTTERDLRTPEBTAAAPKMTKETATTEKTESKITATTTQV 971
Qy 961 TSTTTQDTPPTPKITTLTKTTTLLAIPKVTTKITTTTEIMNKBEETAKPORATNSKATTPK 1020
Db 972 TSTTTQDTPPTPKITTLTKTTTLLAIPKVTTKITTTTEIMNKBEETAKPORATNSKATTPK 1031
Qy 1021 POKPTPAKPKPTSTTKPKTPMVRKPKTPPRKMTSMPELNPSTRJAEMLOTTTBN 1080
Db 1032 POKPTPAKPKPTSTTKPKTPMVRKPKTPPRKMTSMPELNPSTRJAEMLOTTTBN 1091
Qy 1081 QTPNSKLVEVNPKESEDAGAEGETPHMLRPHVEMPEVTPMDVLPVNOGIIINPMLS 1140
Db 1092 QTPNSKLVEVNPKESEDAGAEGETPHMLRPHVEMPEVTPMDVLPVNOGIIINPMLS 1151
Qy 1141 DETNINCKGKPYDGLTTLNGLVAFRGHYFWMLSPSPSPSPARRITTEVWGIPIPIDVFT 1200
Db 1152 DETNINCKGKPYDGLTTLNGLVAFRGHYFWMLSPSPSPSPARRITTEVWGIPIPIDVFT 1211
Qy 1201 RCONCEKTFEPKDSQYMRFTNDIKAGYKPIPKFGGLTGOIVAASTAYKXMBESVY 1260
Db 1212 RCONCEKTFEPKDSQYMRFTNDIKAGYKPIPKFGGLTGOIVAASTAYKXMBESVY 1271
Qy 1261 FFKGGSIOQYIYKQEPVQKCPGRPALNVPYGGEMTOVRRRRFERAIGPSQTHIRIQY 1320
Db 1272 FFKGGSIOQYIYKQEPVQKCPGRPALNVPYGGEMTOVRRRRFERAIGPSQTHIRIQY 1331
Qy 1321 SPARLAYODKGLVHNEVKVSIIMRGLPNUVTSAISLPIRKDGQDYVAFSKDOYNNIDV 1380
Db 1332 SPARLAYODKGLVHNEVKVSIIMRGLPNUVTSAISLPIRKDGQDYVAFSKDOYNNIDV 1391
Qy 1381 PSRTARATTTSGGLSKVWVNCIP 1404
Db 1392 PSRTARATTTSGGLSKVWVNCIP 1415
```

```
KW biodiversity; gene therapy; nutrition.
XX Homo sapiens.
OS MO200154477-A2.
PN 02-AUG-2001.
PD 25-JAN-2001; 2001WO-US02687.
PF 25-JAN-2001; 2000US-0491404.
PR 17-JUL-2000; 2000US-0617746.
PR 03-AUG-2000; 2000US-0631451.
PR 15-SEP-2000; 2000US-0663870.
XX (HYSE-) HYSEQ INC.
PA Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;
PI Cao Y, Drmanac RA, Zhang J, Werhman T;
PI WPI; 2001-476164/51.
DR N-PSDB; AAH98981.
XX Isolated polypeptide for treatment of diseases, diagnostics, raising
PT antibodies and research use -
PT Claim 20; Page 1198-1201; 1275bp; English.
XX The present invention provides the protein and coding sequences of novel
CC proteins from a variety of organisms, including human, dog, cat, horse,
CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea
CC urchin and tomato. These were derived from expressed sequence tags (ESTs)
CC from the organism of interest. They can be used in diagnostics,
CC forensics, gene mapping, identification of mutations, to assess
CC biodiversity and for nutritional purposes. The present sequence is a
CC protein of the invention.
XX
SQ Sequence 1299 AA;
Query Match 92.3%; Score 6950; DB 22; Length 1299;
Best local similarity 100.0%; Pred. No. 0;
Matches 1295; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MAMTLPPIYLILLISVFIQVSSODLSSCAGRGEGYRATNCNCOHWCECPDF 60
Db 1 MAMTLPPIYLILLISVFIQVSSODLSSCAGRGEGYRATNCNCOHWCECPDF 60
Qy 61 KRVTAEISCKGRGCFESFERGREGCDCAQCKYDKCCPDYESFCAEYHNPSPSSKAP 120
Db 61 KRVTAEISCKGRGCFESFERGREGCDCAQCKYDKCCPDYESFCAEYHNPSPSSKAP 120
Qy 121 PPGASQTIKSTYKRSKPKPKKKTKYVISEEITTEHSHVSENQSSSSSSSSSTIW 180
Db 121 PPGASQTIKSTYKRSKPKPKKKTKYVISEEITTEHSHVSENQSSSSSSSSSTIW 180
Qy 181 KIKSSKSAANRELQKLVKNDKNKTKKKRPKPVVVEAGSLGNDGPKVTPPTST 240
Db 181 KIKSSKSAANRELQKLVKNDKNKTKKKRPKPVVVEAGSLGNDGPKVTPPTST 240
Qy 241 TQHNKSTSPKITTAKEINRPSLPNSDTSKETSLTVNKKETVETKETTNNKQSTDG 300
Db 241 TQHNKSTSPKITTAKEINRPSLPNSDTSKETSLTVNKKETVETKETTNNKQSTDG 300
Qy 301 KEKTTSAKETOSIKTSKADLAPTSKVLAKPTPAEFTTGGPALTTKEPTTPKBPAS 360
Db 301 KEKTTSAKETOSIKTSKADLAPTSKVLAKPTPAEFTTGGPALTTKEPTTPKBPAS 360
Qy 361 TTPKEPTTPKSAPTTPKEBPATTPKEBPATTPKEBPATTPKEBPATTPKEBPATTPK 420
Db 361 TTPKEPTTPKSAPTTPKEBPATTPKEBPATTPKEBPATTPKEBPATTPKEBPATTPK 420
Qy 421 APTTPKAPPTPKAPATTPKEBPATTPKEBPATTPKEBPATTPKEBPATTPKEBPATTPK 480
```

```

Db 421 APTTKSAPTTKEBPAPTTPKKAPAPTTKEBPAPTTKEBPTPTTKEBPAPTTKEBPAPTTPK 480
Oy 481 BPAPTAAPKAPAPTTKEBPAPTTKEBPAPTTKEBPTPTTKEBPAPTTKEBPTPTTKEBPAPTT 540
Db 481 BPAPTAAPKAPAPTTKEBPAPTTKEBPAPTTKEBPTPTTKEBPAPTTKEBPTPTTKEBPAPTT 540
Oy 541 TTKSAPTTKEBPTPTTKEBPAPTTKEBPAPTTKEBPAPTTKEBPAPTTKEBPAPTTKEBPAPTT 600
Db 541 TTKSAPTTKEBPTPTTKEBPAPTTKEBPAPTTKEBPAPTTKEBPAPTTKEBPAPTTKEBPAPTT 600
Oy 601 APTAKKEBPAPTTKEBPAPTTKEBPAPTTKEBPAPTTKEBPAPTTKEBPAPTTKEBPAPTTKE 660
Db 601 APTAKKEBPAPTTKEBPAPTTKEBPAPTTKEBPAPTTKEBPAPTTKEBPAPTTKEBPAPTTKE 660
Oy 661 PEEBPAPTTPKAAPPTKEBPAPTTKEBPAPTTKEBPAPTTKEBPAPTTKEBPAPTTKEBPAPTT 720
Db 661 PEEBPAPTTPKAAPPTKEBPAPTTKEBPAPTTKEBPAPTTKEBPAPTTKEBPAPTTKEBPAPTT 720
Oy 721 APTTKKAPKELAPTTTKEBPTSTSDKAPAPTTPKGAPTTAPTTKEBPAPTTKEBPAPTTKEB 780
Db 721 APTTKKAPKELAPTTTKEBPTSTSDKAPAPTTPKGAPTTAPTTKEBPAPTTKEBPAPTTKEB 780
Oy 781 TAPTTKEBPAPTTPKKAPKELAPTTTKEBPTSTSDKAPAPTTPKGAPTTAPTTKEBPAPTTPK 840
Db 781 TAPTTKEBPAPTTPKKAPKELAPTTTKEBPTSTSDKAPAPTTPKGAPTTAPTTKEBPAPTTPK 840
Oy 841 KPAPTTPEBPTTSEVSTPTTKEBPTTHKSPDSTPELSAPTKLENSPKRPGVPT 900
Db 841 KPAPTTPEBPTTSEVSTPTTKEBPTTHKSPDSTPELSAPTKLENSPKRPGVPT 900
Oy 901 TKTPAATKPEMTTAKDKTERDLRTPELTAAAPMTKEATTTTEKTESKITATTTQV 960
Db 901 TKTPAATKPEMTTAKDKTERDLRTPELTAAAPMTKEATTTTEKTESKITATTTQV 960
Oy 961 TSTTTQDTTPEKITTLTKTTTAPKVTYTTTKTITTEINMKPEBIAKPKORATNSKATTPK 1020
Db 961 TSTTTQDTTPEKITTLTKTTTAPKVTYTTTKTITTEINMKPEBIAKPKORATNSKATTPK 1020
Oy 1021 POKPPTAKKPTSTTKKPKTMPRVRCPTTPTPKMTSTMPBELNPTSRIAEMLQTTTTPN 1080
Db 1021 POKPPTAKKPTSTTKKPKTMPRVRCPTTPTPKMTSTMPBELNPTSRIAEMLQTTTTPN 1080
Oy 1081 QTPNSKLEVPNPKSDAGAEGETPHMLLRPHVMEVTPDMDYTPRVNOCIIINPMLS 1140
Db 1081 QTPNSKLEVPNPKSDAGAEGETPHMLLRPHVMEVTPDMDYTPRVNOCIIINPMLS 1140
Oy 1141 DETNINCNGKPVNGLTTLNGLVAFRGHYFWMLSPSPSPPARITTEVWGIPSPIDVFT 1200
Db 1141 DETNINCNGKPVNGLTTLNGLVAFRGHYFWMLSPSPSPPARITTEVWGIPSPIDVFT 1200
Oy 1201 RCNCGKTPFPKDSQYWFNTNDIKDAGYPKPIFKGFGLTGQIVAAALSTAKTKMPEB 1260
Db 1201 RCNCGKTPFPKDSQYWFNTNDIKDAGYPKPIFKGFGLTGQIVAAALSTAKTKMPEB 1260
Oy 1261 FPKRGSIQOYIYKQEPVOKCPGRPALNYPYGE 1295
Db 1261 FPKRGSIQOYIYKQEPVOKCPGRPALNYPYGE 1295

```

```

KM osteopathic.
XX
OS Homo sapiens.
XX
FN WO200064930-A2.
XX
PD 02-NOV-2000.
XX
PF 24-APR-2000; 2000WO-US10953.
XX
PR 23-APR-1999; 99US-0238970.
XX
PA (RHOD-) RHODE ISLAND HOSPITAL LIFESPAN PARTNER.
XX
PI Jay GD;
XX
DR WPI; 2001-024673/03.
XX
PT Novel tribonecetin polypeptide useful as lubricant for treating
XX osteoarthritis, comprises O-linked lubricating moiety
XX
PS Disclosure; Fig 1; 47p; English.
XX
CC The invention relates to a human tribonecetin which is a product of
CC alternative splicing of the human MSF (megakaryocyte stimulating factor)
CC gene. The tribonecetin has at least one O-linked oligosaccharide
CC lubricating moiety and has a polypeptide sequence comprising 1-76
CC repeats of a motif having at least 50% identity to the sequence KEAPPT
CC (AAB29774). The invention also relates to a nucleic acid encoding a
CC human MSF-derived tribonecetin; a biocompatible composition comprising a
CC human tribonecetin for inhibiting tissue adhesion formation; and a method
CC of diagnosing osteoarthritis or a predisposition to osteoarthritis by
CC measuring the amount of MSF or its fragment in a biological sample of a
CC mammal, wherein an increased amount of MSF compared to a control
CC indicates the presence of or predisposition to developing
CC osteoarthritis. The tribonecetin and DNA encoding it are useful in the
CC treatment of osteoarthritis, where they may be used for lubricating
CC mammalian joints, such as articulating joints of humans, dogs or horses.
CC The tribonecetin, when formulated as a membrane, foam, gel or fibre, is
CC useful for inhibiting adhesion between two surfaces such as the injured
CC tissues of a mammal, where the injury is caused by a surgical insertion
CC or trauma, or an artificial device e.g., an orthopaedic implant. In
CC particular, one of the surfaces is pericardial tissue. DNA encoding a
CC tribonecetin may be used in gene therapy. The present sequence represents
CC a substantial portion of a human MSF-derived tribonecetin.
XX
SQ Sequence 902 AA;

```

Query Match 46.3%; Score 3484; DB 22; Length 902;
Best Local Similarity 74.6%; Pred. No. 1.9e-179;
Matches: 794; Conservative 28; Mismatches 80; Indels 162; Gaps 66;

```

Oy 200 VKDNKKORTKKKPKPKPVNDEAGSGLDNGDFKVTTPDTSTQHNKYSTPKITTAKEPN 259
Db 1 VKDNKKORTKKKPKPKPVNDEAGSGLDNGDFKVTTPDTSTQHNKYSTPKITTAKEPN 60
Oy 260 PRPSLPNSTDSTKSTSLTNKETTVEIKETTTNNKQSTDSKERTTSAKTSQSIKTSK 319
Db 61 PRPSLPNSTDSTKSTSLTNKETTVEIKETTTNNKQSTDSKERTTSAKTSQSIKTSK 120
Oy 320 DLAPTSKVLAKPTKAEFTTGGPALTTTKEBPTTTPKEBASTTTPKEBPTTTSAPTTKEBAPT 379
Db 121 DLAPTSKVLAKPTKAEFTTGGPALTTTKEBPTTTPKEBASTTTPKEBPTTTSAPTTKEBAPT 172
Oy 380 EPAPTTTSAPTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPT 439
Db 173 EPAPTTTSAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPT 226
Oy 440 PKRAPPTTKEBPAPTTTKEBPTTTPKEBPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKEB 499
Db 229 -KEBAPT-KEBAPT-KEBAPT-KEBAPT-KEBAPT-KEBAPT-KEBAPT-KEBAPT-KEBAPT-KEB 281
Oy 500 PTTKEBPAPTTTKEBPTTTPKEBPAPTTTTSAPTTTKEBPAPTTTTPKEBPTTKE 559

```

RESULT 6
AAB29778
ID AAB29778 standard; Protein; 902 AA.
AC AAB29778;
XX
XX 28-FEB-2001 (first entry)
XX
XX Human MSF-derived tribonecetin.
XX
XX Human tribonecetin; MSF; megakaryocyte stimulating factor;
XX alternative splicing; joint boundary lubricant; O-linked oligosaccharide;
XX osteoarthritis; tribonecetin; tissue adhesion inhibition;
XX friction coefficient reduction; gene therapy; antiarthritic;

```

Db      282 PTT-KEBAP--TTKEBAPT--KEBAPTKEBAP--TTKEBAPTKEBAPT--KEBAP--TTKE 335
Qy      560 PAPPTKEBAPTTPKAPPTPKBAPTTPKEBAPTTPKAPPTTPKAPAPBAPBAPTTPKEBAPT 619
Db      336 PAPPT--KEBAPT--KEBAPT--KEBAP--TTKEBAPT--KEBAPT--KEBAPT 388
Qy      620 TPKLTPTPEKAPPTPEKAPPTPEBAPTPEBAPTPEBAPTPEBAPTPEBAPTPEBAPTPEBAPT 679
Db      389 T-KEBAPTTPKEP--APTKE--PAPTT--KEBAPT--KEBAPT--KEBAPT--KEBAPT--KE 440
Qy      680 PAPPTKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKE 739
Db      441 PAPPT--KEBAPT--KEBAPT--KEBAPT--KEBAPT--KEBAPT--KEBAPT--KEBAPT--TTK 488
Qy      740 EPSTSDKAPPTPKGAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAP 799
Db      489 EPAPTKE--PAPTT--KEBAPT--KEBAPT--KEBAPT--KEBAPT--KEBAPT--KEBAPT--KE 541
Qy      800 KELAPTTTSGPTSTSDKAPPTPKBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAP 859
Db      542 -----TTKEBAPTTPKE--PAPTT--KEBAPT--KEBAPT--KEBAPTTPKE--PAPTTKEBAP 589
Qy      860 PTTKEBPTTHKSDPSTPBLASAPTPKALNSPKBEGVPTTKTAPATKEBMTTAKDT 919
Db      590 -PTTKEBAPTTPKEBAPT--KEBAP-----TTKEP--APTKEBAPTTPKEBAPTTPKE-- 636
Qy      920 TERDLRTPTPTTAPAPKMTKATATTTEKTESKITATTTOVSTTTODTTPFKITLTKTT 979
Db      637 -----PAPT-----TTKEBAPT--KEBAPTTPKEBAPTTPKEBAPT----- 667
Qy      980 TLAPKTTTKTITTTTEIMNKBEETAKPKORATNSKATTPPKQKPTAPKPKPTSTKKPKT 1039
Db      668 -----TTKEBAPTTPKEP--APT--TKBP-- 686
Qy      1040 MPRVRKPKTTPPKMTSTMBELNPTSRIAEAMLQTTTRPNQTPNSKLVEVNPXSBDAGG 1099
Db      687 -----APTPTPKMTSTMBELNPTSRIAEAML--TTTRPNQTPNSKLVEVNPXSBDAGG 739
Qy      1100 AEGTTPMLLRPHVFMPEVTPDMOYLPRVNOGIIINPMLSDENINCKGKRVDOGLTLRN 1159
Db      740 AEGTTPMLLRPHVFMPEVTPDMOYLPRVNOGIIINPMLSDENINCKGKRVDOGLTLRN 799
Qy      1160 GTLVAFGHYFWMMLSPSPSPSPARITEVWGISPIDVFTRCNCEGTFPFKDSQYWR 1219
Db      800 GTLVAFGHYFWMMLSPSPSPSPARITEVWGISPIDVFTRCNCEGTFPFKDSQYWR 859
Qy      1220 TNDIKDAGYPRKIFKRGGLTGOIUALSTAKYNWPEVYFFK 1263
Db      860 TNDIKDAGYPRKIFKRGGLTGOIUALSTA--YKNWPEVYFFK 902

RESULT 7
ABUS3253
ID ABUS3253 standard; Protein; 551 AA.
XX
AC ABUS3253;
XX
XX 14-APR-2003 (first entry)
XX
OS Homo sapiens.
XX
PN MO200112659-A2.
XX
PD 22-FEB-2001.
XX
PF 18-AUG-2000; 2000MO-IB01496.
XX
PR 18-AUG-1999; 99US-0149499.

```

```

PR      28-SEP-1999; 99US-0156503.
XX
PA (GEHU-) GERMAN HUMAN GENOME PROJECT.
XX
PI Wiemann S;
XX
DR WPI; 2001-327840/34.
XX
PT Nucleic acids having the sequences of clones isolated from libraries of
PT different human tissues, useful in recombinant DNA methodologies -
XX
XX Example III; Page 892-893; 1095bp; English.
XX
PS This invention describes novel polynucleotides and polypeptides isolated
XX from human cDNA libraries which can be used for gene therapy or in
XX vaccines. The polynucleotides of the invention and antibodies encoded by
XX them may be used in the prevention, diagnosis and treatment of diseases
XX associated with inappropriate polypeptide expression. The products of the
XX invention may also be used to identify modulators of expression and
XX activity and to down regulate expression and activity. The antibodies of
XX the invention may also be used as diagnostic agents for detecting the
XX presence of polypeptides in samples. This sequence represents a homologue
XX of a polypeptide described in the disclosure of the invention.
XX
SQ Sequence 551 AA;
XX
Query Match 38.9%; Score 2929; DB 22; Length 551;
Best Local Similarity 100.0%; Pred. No. 8,3e-150; Indels 0; Gaps 0;
Matches 551; Conservative 0; Mismatches 0;
XX
Qy      597 TKKPAPAPKEBAPTTPKEBAPTTPKPLPTPEKLAPTPEBAPTPEBAPTPEBAPTPEBAPT 656
Db      1 TKKPAPAPKEBAPTTPKEBAPTTPKPLPTPEKLAPTPEBAPTPEBAPTPEBAPTPEBAPT 60
Qy      657 TPTPEBAPTTPKAAAPNTKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPT 716
Db      61 TPTPEBAPTTPKAAAPNTKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPT 120
Qy      717 LKEBAPTTPKAPKELAPTTKEPTSTSDKAPPTPKGAPTTPKEBAPTTPKEBAPTTPKEBAPT 776
Db      121 LKEBAPTTPKAPKELAPTTKEPTSTSDKAPPTPKGAPTTPKEBAPTTPKEBAPTTPKEBAPT 180
Qy      777 TPKGAPTTPKEBAPTTPKAPKELAPTTKGPSTSTSDKAPPTPKETAPTTPKEBAPTTPKEBAP 836
Db      181 TPKGAPTTPKEBAPTTPKAPKELAPTTKGPSTSTSDKAPPTPKETAPTTPKEBAPTTPKEBAP 240
Qy      837 TPKKEBAPTTPETPTSTSVSTPTTKEPTTHKSPDESTPBLASAPTPKALNSPKEP 896
Db      241 TPKKEBAPTTPETPTSTSVSTPTTKEPTTHKSPDESTPBLASAPTPKALNSPKEP 300
Qy      897 GVPTTKTPATKPEMTTAKDKTERDLRTPTPTTAAPKMTKEBATTTEKTESKITAT 956
Db      301 GVPTTKTPATKPEMTTAKDKTERDLRTPTPTTAAPKMTKEBATTTEKTESKITAT 360
Qy      957 TTQVSTTTQDTTPFKITLTKTTTLAPKVTTTKKITTITEIMNKBEETAKPKORATNSKA 1016
Db      361 TTQVSTTTQDTTPFKITLTKTTTLAPKVTTTKKITTITEIMNKBEETAKPKORATNSKA 420
Qy      1017 TTPKOKPTAPKPKISTKPKTPMRVRKPKTTPPKMTSTMBELNPTSRIAEAMLQTT 1076
Db      421 TTPKOKPTAPKPKISTKPKTPMRVRKPKTTPPKMTSTMBELNPTSRIAEAMLQTT 480
Qy      1077 TRPNQTPNSKLVEVNPXSBDAGABEGTPMLLRPHVFMPEVTPDMOYLPRVNOGIIIN 1136
Db      481 TRPNQTPNSKLVEVNPXSBDAGABEGTPMLLRPHVFMPEVTPDMOYLPRVNOGIIIN 540
Qy      1137 PMLSDENINCN 1147
Db      541 PMLSDENINCN 551

RESULT 8
ABUS3252

```

ID ABUS3252; standard; Protein; 546 AA.
 AC ABUS3252;
 XX
 DT 14-APR-2003 (first entry)
 DE Human testes-derived DKFzphtes3_4019 homologue #1.
 XX
 KW Human; gene therapy; vaccine; disease treatment; detection.
 OS Homo sapiens.
 XX
 PN WO200112659-A2.
 XX
 PD 22-FEB-2001.
 XX
 PF 18-AUG-2000; 2000WO-1B01496.
 XX
 PR 18-AUG-1999; 99US-0149499.
 PR 28-SEP-1999; 99US-0156503.
 XX
 PA (GERU-) GERMAN HUMAN GENOME PROJECT.
 XX
 PI Wiemann S;
 DR WPI; 2001-327840/34.
 XX
 PT Nucleic acids having the sequences of clones isolated from libraries of
 PT different human tissues, useful in recombinant DNA methodologies -
 PS
 PS Example III; Page 892; 1095pp; English.
 XX
 CC This invention describes novel polynucleotides and polypeptides isolated
 CC from human cDNA libraries which can be used for gene therapy or in
 CC vaccines. The polynucleotides of the invention and antibodies encoded by
 CC them may be used in the prevention, diagnosis and treatment of diseases
 CC associated with inappropriate polypeptide expression. The products of the
 CC invention may also be used to identify modulators of expression and
 CC activity and to down regulate expression and activity. The antibodies of
 CC the invention may also be used as diagnostic agents for detecting the
 CC presence of polypeptides in samples. This sequence represents a homologue
 CC of a polypeptide described in the disclosure of the invention.
 CC
 CC Sequence 546 AA;
 QQ

Query Match	38.8%	Score 299.0	DB 22	Length 546
Best Local Similarity	100.0%	Pred. No. 2.5e-149		
Matched 546	Conservative 0	Mismatches 0	Indels 0	Gaps 0
Qy	488	KKPAPTTPEKAPATTPEKBPATTTKESPTTPEKBPATTTKASAPTTKBPATTTKSAPT	547	
Db	1	KKBPATTTPEKBPATTTPEKBPATTTKESPTTPEKBPATTTKASAPTTTKEBPATTTKSAPT	60	
Qy	548	TPKESPTTTKEBPATTTPEKBPATTTPKCAPATTPEKBPATTTPEKBPATTTKKAPATPKE	607	
Db	61	TPKESPTTTKEBPATTTPEKBPATTTPKCAPATTPEKBPATTTPEKBPATTTKKAPATPKE	120	
Qy	608	PAPTTPEKBPATTTPKKLTPTTPEKAPTTPEKBPATTTPELAPTTPEEPTPTTPEEAPPT	667	
Db	121	PAPTTPEKBPATTTPKKLTPTTPEKAPTTPEKBPATTTPELAPTTPEEPTPTTPEEAPPT	180	
Qy	668	TPKAAAPNTPEKBPATTTPEKBPATTTPEKBPATTTPKGTAPTTTKEBPATTPPK	727	
Db	161	TPKAAAPNTPEKBPATTTPEKBPATTTPEKBPATTTPKGTAPTTTKEBPATTPPK	240	
Qy	728	PAPKELAPTTTKEPSTTSDKAPATTPKGTAPTTPEKBPATTTPEKBPATTTPKGTAPTTLK	787	
Db	241	PAPKELAPTTTKEPSTTSDKAPATTPKGTAPTTPEKBPATTTPEKBPATTTPKGTAPTTLK	300	
Qy	788	BPAPTTPKKBPAPKELAPTTTKEPSTTSDKAPATTPEKBPATTTPEKBPATTTPKKAPATTP	847	
Db	301	BPAPTTPKKBPAPKELAPTTTKEPSTTSDKAPATTPEKBPATTTPEKBPATTTPKKAPATTP	360	

QY	848	TRPPTTSEVSNPTTTEKPTTHKSPDESPELSLAPPTKALENSGKEGVPTTKPAAT	907
Db	361	ETPPPTSEVSNPTTTEKPTTHKSPDESPELSLAPPTKALENSGKEGVPTTKPAAT	420
QY	908	KPEMTTAKDKTERDRLRTPPETTTAAAPMTKEIATTTTEKTESKITTATTTQVSTTTTOD	967
Db	421	KPEMTTAKDKTERDRLRTPPETTTAAAPMTKEIATTTTEKTESKITTATTTQVSTTTTOD	480
QY	968	TTPEKITTLTKTTTLAPKAVTTTKTITTTTEIMNKPEPTAKPKDRAVNSKATTPPKQKPTKA	1020
Db	481	TTPEKITTLTKTTTLAPKAVTTTKTITTTTEIMNKPEPTAKPKDRAVNSKATTPPKQKPTKA	540
QY	1028	PKKPTS 1033	
Db	541	PKKPTS 546	
RESULT 9			
AAOI18634			
ID	AAOI18634	standard; Protein; 538 AA.	
AC	AAOI18634;		
XX			
DT	29-OCT-2002	(first entry)	
XX			
DE	3'	cartilage superficial zone protein coding sequence encoded protein.	
KM	SZP;	superficial zone protein; cartilage; lubrication; human;	
KM	degenerative joint condition; arthritis; osteoporosis; trauma; CACP;		
KM	chondroitin sulphate substitution consensus; antiarthritic;		
KM	antirheumatic; osteopathic; antigout; antiinflammatory; dermatological;		
XX	immunosuppressive.		
OS	Homo sapiens.		
XX			
FH	Key	location/Qualifiers	
FT	Misc-difference 114		
FT	/note= "encoded by ACTACT"		
XX			
PN	WO200262847-A2.		
PD	15-AUG-2002.		
XX			
PF	31-DEC-2001; 2001WO-US50379.		
XX			
PR	29-DEC-2000; 2000US-258920P.		
XX			
PA	(GLAXO) GLAXO GROUP LTD.		
XX	(RUSH-) RUSH PRESBYTERIAN ST LUKE MEDICAL CENT.		
PI	Hutchins JT, Kuetner KE, Schmid TM, Schumacher BL, Su J;		
XX	Dixon EP,		
DR	WPI; 2002-636585/68.		
XX	N-PSDB; AAL49079.		
PT	New purified superficial zone protein (SZP) polypeptides, useful for		
PT	treating degenerative joint conditions, e.g. osteoarthritis, rheumatoid		
PT	arthritis, gout, spondylarthritis, synovitis, tendonitis, lupus, or		
PT	osteoporosis -		
XX			
PS	Claim 59; Page 86-87; 89pp; English.		
CC	The present invention provides the protein and coding sequences of human		
CC	superficial zone protein (SZP). The protein is involved in the		
CC	lubrication of joints, and the sequences can be used in the treatment of		
CC	degenerative joint conditions or to delay symptoms of a degenerative		
CC	joint condition, e.g. osteoarthritis, rheumatoid arthritis, gout,		
CC	psoriatic arthritis, reactive arthritis, viral or post viral arthritis,		
CC	spondylarthritis, juvenile arthritis, synovitis, tendonitis, systemic		
CC	lupus erythematosus, CACP, osteoporosis or trauma. The present sequence		
XX	is the protein encoded by the human 3' cartilage SZP cDNA.		

SQL Sequence 538 AA:
Query Match 37.9%; Score 2850.5; DB 23; Length 538;
Best Local Similarity 99.8%; Pred. No. 1.4e-145;
Matches 538; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
OY 866 PTHHSDESPESAEPTPKALNSPKPEGVPTTKTPAATPKPMPTTAKDTERDUR 925
DB 1 PTHHSDESPESAEPTPKALNSPKPEGVPTTKTPAATPKPMPTTAKDTERDUR 60
OY 926 TTPETTAAPKPKETATTEKTESKITAATTTQVTSSTTQDTPFKITLTKTTAAKV 985
DB 61 TTPETTAAPKPKETATTEKTESKITAATTTQVTSSTTQDTPFKITLTK-TTAAKV 119
OY 986 TTTKTTITTEIMNDEETAKKDRATNSKATTPPKQKPTAKPKPTSTKKPKMPRRK 1045
DB 120 TTTKTTITTEIMNDEETAKKDRATNSKATTPPKQKPTAKPKPTSTKKPKMPRRK 179
OY 1046 PRTTPTPKMTSTMBELNPTSRIAEAMLTQTTTRPNOTPNSKLVENPKSEDAAGAGETP 1105
DB 180 PRTTPTPKMTSTMBELNPTSRIAEAMLTQTTTRPNOTPNSKLVENPKSEDAAGAGETP 239
OY 1106 HMLLRPHVFMPEVTPDMVLPFRVNOGIIINPMLSDEFNI CNKXPVDGLTTLRNGTLVAF 1165
DB 240 HMLLRPHVFMPEVTPDMVLPFRVNOGIIINPMLSDEFNI CNKXPVDGLTTLRNGTLVAF 299
OY 1166 RGHYWMMLSPSPSPSPARITVWGIISPIDVTRFCNCEGTFEFGKSOQVWRFNDIKD 1225
DB 300 RGHYWMMLSPSPSPSPARITVWGIISPIDVTRFCNCEGTFEFGKSOQVWRFNDIKD 359
OY 1226 AGYPKIFKFGGLTGOIIVAAALSTAKYKMPESVYFFKRGGSIOQYIYKQEPVQKPGRR 1285
DB 360 AGYPKIFKFGGLTGOIIVAAALSTAKYKMPESVYFFKRGGSIOQYIYKQEPVQKPGRR 419
OY 1286 PALNTPVYGEWTOVRRRFRERAIIGSQTHTRIRIQSPARLAYODKVLHNEKVSILMRG 1345
DB 420 PALNTPVYGEWTOVRRRFRERAIIGSQTHTRIRIQSPARLAYODKVLHNEKVSILMRG 479
OY 1346 LPNVTSAISLPNTRKPGDYVAASKOYVNIIDPSTARAITTRSGQTLKSWYVNC 1404
DB 480 LPNVTSAISLPNTRKPGDYVAASKOYVNIIDPSTARAITTRSGQTLKSWYVNC 538
RESULT 10
ABUS3254
ID ABUS3254 standard; Protein; 513 AA.
XX AC ABUS3254;
XX DT 14-APR-2003 (first entry)
XX DE Human testes-derived DKFZphc3_4019 homologue #3.
XX KW Human; gene therapy; vaccine; disease treatment; detection.
XX OS Homo sapiens.
XX EN WO200112659-A2.
XX PD 22-FEB-2001.
XX PF 18-AUG-2000; 2000MO-1B01496.
XX PR 18-AUG-1999; 99US-0149499.
XX PR 28-SEP-1999; 99US-0156503.
XX PA (GEHU-) GERMAN HUMAN GENOME PROJECT.
XX PI Wiemann S;
XX DR WPI; 2001-327840/34.
XX PT Nucleic acids having the sequences of clones isolated from libraries of

PT different human tissues, useful in recombinant DNA methodologies -
XX Example III; Page 893; 1095pp; English.
XX CC This invention describes novel polynucleotides and polypeptides isolated
XX CC from human cDNA libraries which can be used for gene therapy or in
XX CC vaccines. The polynucleotides of the invention and antibodies encoded by
XX CC them may be used in the prevention, diagnosis and treatment of diseases
XX CC associated with inappropriate polypeptide expression. The products of the
XX CC invention may also be used to identify modulators of expression and
XX CC activity and to down regulate expression and activity. The antibodies of
XX CC the invention may also be used as diagnostic agents for detecting the
XX CC presence of polypeptides in samples. This sequence represents a homologue
XX CC of a polypeptide described in the disclosure of the invention.
SQL Sequence 513 AA:
Query Match 36.6%; Score 2757; DB 22; Length 513;
Best Local Similarity 100.0%; Pred. No. 1.4e-140;
Matches 513; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 207 RTKKKPTPKPVVDAGSGLDNGDFKTTPTSTTQHNKSTSKITTAKNRPSILP 266
DB 1 RTKKKPTPKPVVDAGSGLDNGDFKTTPTSTTQHNKSTSKITTAKNRPSILP 60
OY 267 NSDTSKETSILVNKEETVETKETTNTKNTSDGKEKTSKETSKEKTSKXDLAPTSK 326
DB 61 NSDTSKETSILVNKEETVETKETTNTKNTSDGKEKTSKETSKEKTSKXDLAPTSK 120
OY 327 VLAAPTPKAEETTTGPAALTPPKBPTTPPKBAPTTPPKBAPTTPPKBAPTTPPKBAPTTP 386
DB 121 VLAAPTPKAEETTTGPAALTPPKBPTTPPKBAPTTPPKBAPTTPPKBAPTTPPKBAPTTP 180
OY 387 KSAPTTPKBPATTPPKBAPTTPPKBAPTTPPKBAPTTPPKBAPTTPPKBAPTTPPKBAPT 446
DB 181 KSAPTTPKBPATTPPKBAPTTPPKBAPTTPPKBAPTTPPKBAPTTPPKBAPTTPPKBAPT 240
OY 447 TPKEBAPTTPKEBAPTTPPKBAPTTPPKBAPTTPPKBAPTTPPKBAPTTPPKBAPTTPKEP 506
DB 241 TPKEBAPTTPKEBAPTTPPKBAPTTPPKBAPTTPPKBAPTTPPKBAPTTPPKBAPTTPKEP 300
OY 507 APPTTKESPPTPKBAPTTPPKBAPTTPPKBAPTTPPKBAPTTPPKBAPTTPPKBAPTTPK 566
DB 301 APPTTKESPPTPKBAPTTPPKBAPTTPPKBAPTTPPKBAPTTPPKBAPTTPPKBAPTTPK 360
OY 567 EPAPTTPKBPATTPPKBAPTTPPKBAPTTPPKBAPTTPPKBAPTTPPKBAPTTPPKBAPT 626
DB 361 EPAPTTPKBPATTPPKBAPTTPPKBAPTTPPKBAPTTPPKBAPTTPPKBAPTTPPKBAPT 420
OY 627 TTPKLAAPTPKEKAPATTPPEELAPTTPEEBPTTPPEBAPTTPPKAANPTPKBAPTTPPK 686
DB 421 TTPKLAAPTPKEKAPATTPPEELAPTTPEEBPTTPPEBAPTTPPKAANPTPKBAPTTPPK 480
OY 687 EPAPTTPKBPATTPPKBAPTTPPKBAPTTPPKBAPTTPPKBAPTTPPKBAPTTPPKBAPT 719
DB 481 EPAPTTPKBPATTPPKBAPTTPPKBAPTTPPKBAPTTPPKBAPTTPPKBAPTTPPKBAPT 513
RESULT 11
AAR80041
ID AAR80041 standard; Protein; 452 AA.
XX AC AAR80041;
XX DT 25-MAR-2003 (updated)
XX DT 10-APR-1996 (first entry)
XX DE Human megakaryocytopoietin protein.
XX KW Human; megakaryocytopoietin; wheat germ agglutinin; heparin;
XX KW megakaryocyte; aplastic anaemia; thrombocytopenia; hemacytopenia;
XX KW multipotential stem cell.


```

OS Homo sapiens.
XX Key Location/Qualifiers
FH MISC-difference 393..396
FT MISC-difference /note="unspecified amino acids"
FT MISC-difference 444..446
FT MISC-difference /note="unspecified amino acids"
XX MO9523861-A1.
XX
XX PD 08-SEP-1995.
XX
XX PR 06-MAR-1995; 95MO-CN00015.
XX
XX PR 04-MAR-1994; 94CN-0112066.
XX
XX (SHAN-) SHANGHAI BEITE BIOTECHNOLOGY CO LTD.
XX
XX PI Gu X, Han Z, Shen Q;
XX
XX DR MPI; 1995-320576/41.
XX DR N-PSDB; AAT04546.
XX
XX PT New haematopoietic cell growth factor - used for treating
XX thrombocytopenia and hematocytopenia
XX
XX PS Example; Page 23; 36pp; Chinese.
XX
XX CC This sequence represents the human megakaryocytopoietin (MPO) protein.
XX CC This sequence was purified using a carrier which can couple wheat germ
XX CC agglutinin and heparin to separate MPO. Fragments of this sequence (see
XX CC AAR80039 and AAR80040) were used to produce the amplification primers
XX CC shown in AAT04544 and AAT04545. The fragments amplified by these primers
XX CC can then be used as probes to screen human cDNA libraries for MPO cDNA.
XX CC The MPO cDNA can then be inserted into a plasmid which is used to
XX CC transform cells to produce MPO. The MPO sequence is capable of promoting
XX CC colony formation of megakaryocytes, enlarging the size of megakaryocytes
XX CC and stimulating the proliferation of multipotential stem cells. The
XX CC factor may be used for treating thrombocytopenia and hematocytopenia.
XX CC The purification method can be used to isolate MPO from human urine or
XX CC serum of patients with aplastic anaemia, and from animal blood or urine
XX CC by radiation exposing the animals to induce aplastic anaemia.
XX CC (Updated on 25-MAR-2003 to correct PA field.)
XX
XX SQ Sequence 452 AA;

```

```

QY 361 TTPKEPTTTIKSAPTPKEPAPTTTKSAPTTPKEPAPTTTKEPAPTTKEPAPTTTKSP 420
DB 318 TTPKEPTTTIKSAPTPKEPAPTTTKSAPTTPKEPAPTTTKEPAPTTKEPAPTTTKSP 377
QY 421 APTTTKSAPTPKEPAPTTPKPAPTTPKEPAPTTKEPAPTTKEPAPTTKEPAPTTKEPA 480
DB 378 APTTTKSHPLPLRSCXXXCTGP-----TPKEPHPLPLRSHPLPTKEPAPTTKE 426
QY 481 EPAPTAPKKPAPTTKE 497
DB 427 EPAPTAPKKPAPLPLE 443

```

RESULT 12

```

AAB60569
ID AAB60569 standard; Protein; 472 AA.
XX
XX AC AAB60569;
XX
XX DT 27-APR-2001 (first entry)
XX
XX DE Bovine MSF orthologue, superficial zone protein (SZP).
XX
XX KM Bovine; CACP protein; campodactylly-arthropathy-coxa vara-pericarditis;
XX KM superficial zone protein; SZP; MSF orthologue; synovial lubricant;
XX KM osteoarthritis; joint lubrication; osteopathic; antiarthritic.
XX
XX OS Bos taurus.
XX
XX PN WO200107068-A1.
XX
XX PD 01-FEB-2001.
XX
XX PF 21-JUL-2000; 2000WO-US20002.
XX
XX PR 23-JUL-1999; 99US-0145328.
XX PR 19-JUL-2000; 2000US-0145328.
XX
XX PA (UYCA-) UNIV CASE WESTERN RESERVE.
XX
XX PI Warman ML;
XX
XX DR MPI; 2001-182721/18.
XX
XX PT New composition comprising the campodactylly-arthropathy-coxa
XX PT vara-pericarditis protein in combination with an anesthetic, useful for
XX PT treating osteoarthritis, or as lubricants of tissue and joints
XX
XX PS Example 1; Fig 4; 34pp; English.
XX
XX CC The invention relates to a method of treating osteoarthritis via the
XX CC administration of a composition comprising the campodactylly-arthropathy-
XX CC coxa vara-pericarditis (CACP) protein, or portions of the CACP protein.
XX CC The composition may further comprise a local anesthetic. The composition
XX CC of the invention may be administered via intra-articular or intravenous
XX CC injection. The human CACP protein is identified in the invention as
XX CC being megakaryocyte stimulating factor (MSF). The gene encoding
XX CC CACP protein (MSF) is located on chromosome 1q25-31, and mutations in
XX CC this gene are responsible for the heritable disorder campodactylly-
XX CC arthropathy-coxa vara-pericarditis in which patients have synovial
XX CC hyperplasia without evidence of inflammation. CACP protein (MSF)
XX CC acts as a synovial lubricant, and can be used to lubricate tissue and
XX CC joints in the treatment of osteoarthritis. The composition may be
XX CC applied to reduce the symptoms of osteoarthritis (e.g., joint pain,
XX CC loss of range of movement or joint damage). The present sequence
XX CC represents the bovine orthologue of human MSF, superficial zone
XX CC protein (SZP).
XX
XX SQ Sequence 472 AA;

```

Query Match 22.7%; Score 1707.5; DB 22; Length 472;
Best Local Similarity 79.9%; Pred. No. 3.8e-84;

[illegible]

RESULT 13
AAU11261
ID AAU11261 standard; Protein; 292 AA

Seq	Sequence	292 AA:	20.5%;	Score 1545;	DB 23;	Length 292;
	Query Match		86.9%;	Pred. No. 1.2e-75;		
	Best Local Similarity					
	Matches 292;	Conservative 0;	Mismatches 0;	Indels 44;	Gaps 1	
QY	41	DATCNCDYNCQHWNECCEDPFKRVCTAEIASCRCRCESPPRGHECTCDACCKYDKCCPY	100			
DB	1	DATCNCDYNCQHWNECCEDPFKRVCTAEIASCRCRCESFEREGECDDACCKYDKCCPY	60			
QY	101	ESFCAEVANNPSPSSSKAPPPSGASQITIKSTTKSPRPNNKKTKKVIRESREITMEHSV	160			
DB	61	ESFCAEVANNPSPSSSKAPPPSGASQITIKSTTKSPRPNNKKTKKVIRESREITE----	116			
QY	161	SENQSSSSSSSSSSSTIIWKIKSSKNSAANRELQKLVKNNKNNRTKKKRLPPRPVVD	220			
DB	117	-----KNNKNNRTKKKRLPPRPVVD	136			
QY	221	EAGSGLDNGDFVYITPDTSTTOHNVSTSPKITTKAPINPRSLPPNSDTSKETSITVNN	280			
DB	137	EAGSGLDNGDFVYITPDTSTTOHNVSTSPKITTKAPINPRSLPPNSDTSKETSITVNN	196			
QY	281	ETTVAETKETTNNKOSTDGEKKTSAKETOSIEKTSADKLAPTSKVLAKPPKAEYTTK	340			
DB	197	ETVEIKETITNNKOSTDGEKKTSAKETOSIEKTSADKLAPTSKVLAKPPKAEYTTK	256			
QY	341	GPALTTPKEPTPTTTPKEBPASTTPKEPTPTTIKSAPT	376			
DB	257	GPALTTPKEPTPTTTPKEBPASTTPKEPTPTTIKSAPT	292			

RESULT 14
AAM24516
ID AAM24516 standard; Protein; 5179 AA.

XX AAU11261;
AC 12-MAR-2002 (first entry)
DT
XX
XX Human HAPO polypeptide.
DE
XX Human; HAPO; stimulating factor; haemopoietic stem ancestral cell;
KW vascular endothelial cell; haemopoietic disease; vascular disease.
XX
XX Homo sapiens.
OS
XX CNJ312294-A.
PN
XX 12-SEP-2001.
PD
XX 28-FEB-2001; 2001CN-0109083.
PF
XX 28-FEB-2001; 2001CN-0109083.
PR
XX (HEMA-) INST HEMATOLOGY CHINESE ACAD MEDICAL SCI.
PA
XX Han Z, Liu Y, Cai Y;
PI
XX WPI; 2002-018506/03.
DR
XX N-PSDB; AAS16924.
DR
XX Human blood and blood vessel cytopoiesis hormone and its preparation -
PT
XX
PS Claim 1; Page 9 (Disclosure); 20pp; Chinese.
XX
XX The invention relates to a stimulating factor, HAPO, with the effect of
CC promoting growth of haemopoietic stem ancestral cells and vascular
CC endothelial cells. HAPO is characterised using several pairs of primers
CC specific for HAPO DNA to make a series of gene segments by PCR
CC amplification. An antibody specific to the HAPO polypeptide can be used
CC in the treatment and prevention of haemopoietic and vascular diseases.
XX This sequence represents the HAPO polypeptide of the invention.

XX	PAM24516;
AC	
XX	
DT	12-OCT-2001 (first entry)
XX	
DE	C899P predicted amino acid sequence.
XX	
KW	Human; immunotherapy; diagnosis; colon cancer; colon tumour;
XX	immunogenic; gene therapy; vaccine; colonic cancer.
OS	Homo sapiens.
XX	
PN	WO200149716-A2.
XX	
PD	12-JUL-2001.
XX	
PF	29-DEC-2000; 2000WO-US35596.
XX	
PR	30-DEC-1999; 99US-0476296.
PR	10-JAN-2000; 2000US-0480321.
PR	15-FEB-2000; 2000US-0504629.
PR	06-MAR-2000; 2000US-0519444.
PR	19-MAY-2000; 2000US-0575251.
PR	29-JUN-2000; 2000US-0609448.
PR	28-AUG-2000; 2000US-0649811.
XX	
PA	(CORI-) CORIXA CORP.
XX	
PI	Xu J, Lodes MJ, Secrist H, Benson DR, Meagher MJ, Stolk JA;
PI	King GE, Wang T, JIang Y;
XX	
DR	WPI, 2001-441847/47.
XX	
PT	Colon tumor associated proteins and nucleic acids useful for the
XX	prevention, diagnosis and treatment of colonic cancer -
PS	Claim 2; Page 446-462; 472pp; English.
CC	The present invention describes colon tumour associated proteins (I) and

CC the polynucleotides (II) that encode them. (I) have cytosolic activity.
 CC (I) and (II) can be used in gene therapy and vaccine production. (II) and
 CC (II) may be used in the prevention, diagnosis and treatment of diseases
 CC associated with inappropriate colon tumour associated protein (TCAP)
 CC expression, such as colorectal cancer. For example, (I) and (II) may be
 CC used to treat disorders associated with decreased expression by
 CC rectifying mutations or deletions in a patient's genome that affect the
 CC activity of TCAPs by expressing inactive proteins or to supplement the
 CC patient's own production of them. Additionally, (II) may be used to
 CC produce the TCAP proteins, by inserting the nucleic acids into a host
 CC cell culturing the cell to express the protein. (II) and its
 CC complementary sequences may also be used as DNA probes in diagnostic
 CC polymerase chain reaction (PCR) and hybridisation assays to detect and
 CC quantitate the presence of similar nucleic acids in samples, and
 CC therefore which patients may be in need of restorative therapy. (I) may
 CC also be used as antigens in the production of antibodies against TCAPs
 CC and in assays to identify modulators of TCAP expression and activity.
 CC Anti-(I) antibodies and antagonists may also be used to down regulate
 CC TCAP expression and activity. The anti-(I) antibodies may also be used
 CC as diagnostic agents for detecting the presence of TCAPs in samples
 CC (e.g. by enzyme linked immunosorbent assay (ELISA)). AA128460 to AA129512
 CC and AA24494 to AA24523 represent nucleotide and amino acid sequences
 CC given in the exemplification of the present invention.

XX Sequence 5179 AA;

Query Match 15 8%; Score 1188; DB 22; Length 5179;

Best Local Similarity 30.9%; Pred. No. 4.7e-55; Indels 279; Gaps 54;

Matches 405; Conservative 87; Mismatches 539;

4 KTLPIYLLLLSVFIQVVSQDLSSC--AGRCG-----EGYSRDATCNDYNGHYMC 56
 1169 KDRPIY-----EEDLKVCVADKCGCVDETHYPRGASVTEETCSGC-V 1213
 57 CPDFKVCYTAELSCGRCFSEFEG-----RECDCAQCKYDKC-----PDYSFCME 106
 1214 TNSGVVCPRE--EGKINOTODGAFCEICGPGVTEKHPNICSITRSTLTFTT 1270
 107 VNPPTSPS-SKKAPPSGASOTIKSTRSPKPKKTKKVISEETEEH--SVSN 163
 1271 IYLPPTTPTSTTTTPTSTVSTLT--PK-----LCCLMSDWNEHPSGSD 1320
 164 QSSSSSSSSSSSTIMKISSKSNANR-----LQKLKYDNKNTKKKPTPKPV 218
 1321 GDRPPDGVCGAPBDI-ECRSYKDPHLSGEGQKQCVSVGFIKME----- 1368
 219 VDEAGSGLDN--GDFKVTTPDTSTQHNKYSTSPK-ITTAKPINRPSLPPNSDTSKETS 275
 1369 -DQFGNGPFGLCYDYKIRV-----NCCWPMDCITTPSPPTTTPSPPTTTTLPPT 1419
 276 LTVNKETVETKETTNNKQSTDGKETSATAKIOSIKTSAKDLAPSKVLAKPTPKA 335
 1420 TTPSPPTT-----TTTTPTTTPSPPTTTTTP-----LPTT-----TPSPPI 1458
 336 ETTTGKPAITPKKEPT-----PTPKEPASTPKKEPTTTPKAP--TPKKEP-----A 382
 1455 STTTTTPPTTTPSPPTTTPSPPTTTPPTTTPPTTTPSPPTTTPPTTTPPTTTPPT 1518
 383 PTTTKAP--TPPKAPATTKEAPATTKEP-----APTTKBAPATTTKAP--TTP 432
 1519 PTTTSPPTTTPPTTTPPTTTPSPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPT 1578
 433 KEAPATTPKKAPATTPKKAPATTPKKAPATTPKKAPATTPKKAPATTPKKAPATT 485
 1579 SPPTTTPSPPTTTPPTTTPSPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPT 1638
 486 APKAPATTPKKAPATTPKKAPATTPKKAPATTPKKAPATTPKKAPATTPKKAPATT 545
 1639 TTPSPPTTTPPTTTPPTTTPSPPTTTPPTTTPPTTTPPTTTPSPPTTTPPTTTPPT 1698
 546 P--TTPKESPPTTKEAPATTPKKAPATTPKKAPATTPKKAPATTPKKAPATT 602
 1699 PPTTTPSPPI-TTTTTPSTTTPSPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPT 1757

QY 603 TAPKEAPATTTP-----KEAPATTTP----- 621
 DB 1758 TTPLEPSTTPPTPSPPTTTPPTTTPCVPLCNMTGMLDSGKNFHKRGDTLIGDVGCGW 1817
 QY 622 -----KKLP----- 626
 DB 1818 AANISCRATMPDVPDIGLGQTVCDVSGLICKNEQKPGVIMAFCLANINEVQCE 1877
 QY 627 --TTPKEAPATTTPKEAPATTPEELAPTTPEEPTTP-TPBBAPATTPKAAPTPKEAPT 683
 DB 1878 CVTQPTMTTTPTEENPTTTPPTTTPPTTTPVTPPTPGTQPTTTPPTTTPVTPPTPT 1937
 QY 684 TPKEP-----APTKEKAPATTTPKEKAPATTTPKGAAPTTLKEAPATTKEKAPKLAAPT 737
 DB 1938 GTQPTTTPPTTTPPTTTPVTPPTPGTQPTTTPPTTTPVTPPTPGTQPTTTPPTTTPPT 1997
 QY 738 TKEPTSTSDKAP--TPKGAAPTTPKKAPATTTPKGAAPTTLKEAPATTPEK 796
 DB 1998 TTPPTPTPTGQPTTTPPTTTPPTTTPVTPPTPGTQPTTTPPTTTPVTPPTPGTQPT 2056
 QY 797 PAPKELAPTTKGPSTSTSDKAP--TPKETAAPTTPKKAPATTPEKAPATTPEPTPTS 855
 DB 2057 PTTTPTTTPPTTTPPTTTPPTPGTQPTTTPPTTTPVTPPTPGTQPTTTPPTTTPPT 2115
 QY 856 EVSTPTTKEPTTHKSPDSSTPELSAAPTKEALENSKEP-----GVPT-TKTP 904
 DB 2116 PTPPTPGTQPTT--TPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPT 2172
 QY 905 AATKEPTTAKDKTERDLR--TTP--ETTAAPKMT--KETATTEKTESKITAT 956
 DB 2173 TTTPTTTPPTTTPPTTTPPTPGTQPTTTPPTTTPVTPPTPGTQPTTTPPTTTPPT 2232
 QY 957 TTQVSTTTPQTPPKITTLKTTLAPKVTTP-KKITTTEIMNKEEPTANPKDRATNSK 1015
 DB 2233 PTP--TGTPPTTTP--ITT--TTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPT 2287
 QY 1016 AATPKPKATKAP--KKPTSTKKPKTPVRKPKTTPTRKMTSNMELAN--TSRIKIA 1071
 DB 2288 TTTPTTTPPTTTPPTTTPPTPGTQPTTTPPTTTPVTPPTPGTQPTTTPPTTTPPT 2347
 QY 1072 MLOQTTT--PNOTPNSKLEVNPKSEDAGAGETPHMLRPHVEMEPTP 1120
 DB 2348 PTPPTQPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPT 2392

RESULT 15
 ABP55365
 ID ABP55365 standard; Protein, 5179 AA.
 XX
 AC ABP55365;
 XX
 DT 30-JAN-2003 (first entry)
 XX
 DE Human colon tumour protein for clone C899P SEQ ID NO:1068.
 XX
 KM Human; colon cancer; colon tumour; immunotherapy; diagnosis; cancer;
 KM tumour; immune response; immunostimulant; cytostatic; vaccine.
 OS Homo sapiens.
 PN MO200283070-A2.
 PD 24-OCT-2002.
 PF 09-APR-2002; 2002MO-US11475.
 PR 10-APR-2001; 2001US-0833263.
 PR 03-AUG-2001; 2001US-0922217.
 PR 19-DEC-2001; 2001US-0025380.
 XX
 XX (CORI-) CORIXA CORP.
 XX

PI Xu J, Lodes MJ, Secriet H, Benson DR, Meagher MJ, Stolk JA;
PI Wang T, Jiang Y, Smith CL, King GR, Wang A, Clapper JD;
PI Skeiky YAM, Fanger GR, Vedavick TS, Carter D;
XX WPI; 2003-067548/06.
DR N-PSDBJ; AB233690.
XX
PT New polynucleotide, useful for the preparation of a composition for
PT stimulating an immune response against, or treating, cancer -
XX
PS Disclosure; Page 447-464; 537pp; English.
XX
CC The present invention describes compound (1) for the immunotherapy and
CC diagnosis of colon cancer. Also described: (1) a method for detecting and/or
CC the presence of cancer in a patient; (2) a method for stimulating and/or
CC expanding T cells specific for a tumour protein; (3) an isolated T cell
CC population comprising T cells prepared by the method of (2); (4) a method
CC for stimulating an immune response in a patient; (5) a method for
CC treating cancer in a patient; and (6) a method for inhibiting the
CC development of cancer in a patient. (1) have immunostimulant and
CC cytotoxic activities and can be used in vaccines. AB232646 to AB233725
CC and AB25343 to AB25391 represent human colon cancer/tumour related
CC sequences used in the exemplification of the present invention.
XX
SQ Sequence 5179 AA;

Query Match 15.8%; Score 1188; DB 24; Length 5179;
Best Local Similarity 30.9%; Pred. No. 4.7e-55;
Matches 405; Conservative 87; Mismatches 539; Indels 279; Gaps 54;

QY 4 KTLPLYLILLVFIQGVSSODLSSC--AGRCG-----EGYSRATGNCNDYQHWEC 56
DB 1169 KDRPLY-----BEDLKKCVYADKCGCYVEDTHYPASVPTETSCV-C 1213
QY 57 CPDFKVCVTAELSCRCFESEFEG-----RECDCAQCKKYDKC---PDYSEFCAE 106
DB 1214 TNSQVCRPE--EGKILNQDGAFCYWEICGNGVKEKFNCSITTRPSTLTFTT 1270
QY 107 VANPSPSPS-SKKAPPSGASOTISTKSPKPKKKTKKVISEITEBH--SVEN 163
DB 1271 ITLPTTPSTFTTTTTTPTSTVLTST--PK-----LCCLMSDWINEBHPSSGDD 1320
QY 164 QESSSSSSSSSSSTIMKISKNSAARE-----LOKKLKYDKQKQRTKKKPPKPV 218
DB 1321 GDRPFDGVCAPEDI-ECRSYKDPHLSHQGVQCDVSGVFCXBE----- 1368
QY 219 VDEAGSGLDN--GDPKVTTPDSTTQHNVKSTSPK-ITTAKEINRPSLPPNSDTSKETS 275
DB 1369 -DQFGNGPFGLCYDKIRV-----NCCWPMDCITTPSPPTTTPSPPTTTTLPPT 1419
QY 276 LTVNKEITVEIKETTTNKQSTDGKEKTSKETSQISAKDLAPTSKVLAKPTPKA 335
DB 1420 TTPSPPTT-----TTTTPTTPPTSPPIITTTTP-----LPTT-----TPSPPI 1458
QY 336 ETTTGPALTPKEPT-----PTPKBASTPKKEPTTIKSAF--TTPKEP-----A 382
DB 1459 STTTPPTTPSPPTTTPSPPTTTPSPPTTTPPTTTPSPPTTTPPTTTPPTTTP 1518
QY 383 PTTTSAF--TTPKEAPTTTKEAPTTKEP-----APTTPKEAPTTTTSAP--TTP 432
DB 1519 PTTTSPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTP 1578
QY 433 KEPAFTTPKKAFTTPKKAFTTPKKEPTTTPKKAFTTPKKAFTTPKKEP-----APT 485
DB 1579 SPPTTTPSPPTTTPPTTTPSPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPT 1638
QY 486 APKKAFTTPKKAFTTPKKAFTTPKKAFTTPKKAFTTPKKAFTTPKKAFTTPKKA 545
DB 1639 TTPSPPTTTPPTTTPSPPTTTPSPPTTTPPTTTPSPPTTTPSPPTTTPPTTTPMT 1698
QY 546 P---TTPKEPSPTTTPKKAFTTPKKAFTTPKKAFTTPKKAFTTPKKAFTTPKKA 602
DB 1699 PSPPTTTPSPPTTTPPTTTPSPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPT 1757

QY 603 TAPKKAFTTP-----KETATTP----- 621
DB 1758 TTPLPSPITTPPTSPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPT 1817
QY 622 -----KKLTP----- 626
DB 1818 AANISCRATWPDVPIQLOGTVCVSVGLCKNEDQKRGVAPMAFCLNVEINVOCE 1877
QY 627 --TTPKLAFTTPKKAFTTPPEELAPTTPEEPPTT--TPEBAPTTPKAAANTPKAFT 683
DB 1878 CVTQPTTMTTTTTTNNPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPT 1937
QY 684 TPKRP-----APTTPKBAFT--TPKETAFTTPKGTATPTLKEBAPTTPKKAPELAPT 737
DB 1938 GTQPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPT 1997
QY 738 TKEPTSTSDKPAF--TTPKGAFTTPKKAFTTPKKAFTTPKKAFTTPKKAFTTPKKA 796
DB 1998 TTPPTTPPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPT 2056
QY 797 PAPKELAPTTTGTSTSDKPAF--TTPKETAFTTPKKAFTTPKKAFTTPKKAFTTPPT 855
DB 2057 PTT 2115
QY 856 EVSTPTTTPKPTTTHKSPDSSTPBLASAPTPKALENSPKP-----GVPT--TKTP 904
DB 2116 PTPPTTGTQPTT--TPITTTTTPPTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPT 2172
QY 905 AATPEMTTAKDTERDLR--TTP--ETTAAAPMT--KETATTEKTESKITAT 956
DB 2173 TTTPTT 2232
QY 957 TTQVSTTTQDTPPKITTLKTTTLAPKVTT--KTIITTEINAKPEBIAKPAKRAINSK 1015
DB 2233 PTP--TGQTPPTTP--ITT--TTTTPPTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTT 2287
QY 1016 ATEPKQPKTPAP--KKPTSTKPKTPMRYKPKTTPPKMTSTMPELN--TSRIABA 1071
DB 2288 TTTPTT 2347
QY 1072 MLQTTTR--PNOGPSKLVENPKSEDAGAGETPHMLBPHVMEPTV 1120
DB 2348 PTPPTT 2392

Search completed: December 8, 2003, 09:35:14
Job time : 59 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 8, 2003, 09:31:27 ; Search time 53 Seconds

(without alignments)
6835.960 Million cell updates/sec

Title: US-09-556-246-1

Perfect score: 7526
Sequence: 1 MAMKTLPIVILLLSVFIQ.....ARATRRSGQTLSKWNCP 1404

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

1: SP archaea:*
2: SP bacteria:*
3: SP fungi:*
4: SP human:*
5: SP invertebrate:*
6: SP mammal:*
7: SP mhc:*
8: SP organelle:*
9: SP phage:*
10: SP plant:*
11: SP rodent:*
12: SP virus:*
13: SP vertebrate:*
14: SP unclassified:*
15: SP yvirus:*
16: SP bacteriaph:*
17: SP archaep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7526	100.0	1404	4	Q92954
2	7497	99.6	1404	4	Q9BX49
3	3968	53.0	1054	11	Q9JW99
4	1713.5	22.8	401	6	077765
5	1215	16.1	1152	5	Q9VR49
6	982.5	13.1	1079	5	Q9N4S7
7	974	12.9	3458	5	Q8IR52
8	968.5	12.9	3432	5	Q8IR51
9	902	12.0	1795	5	076894
10	896.5	11.9	1349	4	Q8MWQ4
11	863.5	11.5	1489	10	Q96449
12	846	11.2	1274	5	Q20007
13	843.5	11.2	934	5	Q9VVG2
14	829	11.0	1607	10	Q8H6Q5
15	816	10.8	1315	10	Q9SPM0
16	812	10.8	2187	11	P70670

17	811.5	10.8	1480	10	Q9LIE8
18	784	10.4	1188	10	Q41805
19	782.5	10.4	2112	5	Q9VEL9
20	767.5	10.2	1009	10	Q8L685
21	746	9.9	1254	5	Q94185
22	715	9.5	763	2	Q9XDH2
23	710	9.4	4969	11	Q8CF91
24	710	9.4	5165	11	Q8CF92
25	709.5	9.4	18519	5	Q8ISF6
26	709.5	9.4	18534	5	Q8ISF7
27	703	9.3	10578	5	Q8ISF5
28	695	9.2	1151	13	Q57580
29	678	9.0	2090	4	Q96QC2
30	677	9.0	598	16	Q8VKN7
31	677	9.0	2114	5	Q8I018
32	675.5	9.0	2768	5	Q9VC00
33	674.5	9.0	2089	4	Q14676
34	669.5	8.9	2284	5	Q9VPG1
35	666.5	8.9	7962	4	Q10465
36	665.5	8.8	34350	4	Q8WZ42
37	643.5	8.6	990	13	Q91803
38	636	8.5	6632	5	Q17362
39	635	8.4	1720	5	Q81486
40	633	8.4	489	10	Q41707
41	632	8.4	761	10	Q92010
42	631	8.4	2344	5	Q9N3Y8
43	629	8.4	6632	5	Q01761
44	627.5	8.3	971	5	Q9XV54
45	622.5	8.3	839	16	Q9RX57

ALIGNMENTS

RESULT 1
Q92954 PRELIMINARY; PRT; 1404 AA.
ID Q92954
AC Q92954;
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DE 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Megakaryocyte stimulating factor.
OC Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_Taxid=9606;
RX [1]
RP SEQUENCE FROM N.A.
RA Turner K.J., Fitz L.J., Temple P., Jacobs K., Larson D., Leary A.C.,
RA Kelleher K., Giannotti J., Calvetti J., Fitzgerald M., Kriz M.J.,
RA Ferenz C., Grobholz J., Fraser H., Bean K., Norton C.R., Gesner T.,
RA Bhattacharya S., Kriz R., Hewick R., Clark S.C.;
RT "Purification, Biochemical Characterization, and Cloning of a Novel
RT Megakaryocyte Stimulating Factor that has Megakaryocyte Colony
RT Stimulating Activity."
RL Blood 78:279-279(1991).
RN [2]
RP SEQUENCE FROM N.A.
RA Werberg D.M., Fitz L.J., Temple P., Giannotti J., Murtha P.,
RA Fitzgerald M., Scaltreco J., Kelleher K., Plesner K., Kriz R.,
RA Jacobs K., Turner K.;
RT "A Comparison of Vitronectin and Megakaryocyte Stimulating Factor."
RL (In) Plesner K.T., Rosenblatt S., Kost C., Wegernoff J.,
RL Mosher D.F. (eds.);
RL BIOLOGY OF VITRONECTINS AND THEIR RECEPTORS., pp.45-52,
RL Elsevier Science Publishers B.V. (1993).
RN [3]
RP SEQUENCE FROM N.A.
RA Turner K.J., Fitz L.J., Temple P., Jacobs K., Larson D., Leary A.C.,
RA Kelleher K., Giannotti J., Calvetti J., Fitzgerald M., Kriz M.J.,
RA Ferenz C., Grobholz J., Fraser H., Bean K., Norton C.R., Gesner T.,
RA Bhattacharya S., Kriz R., Hewick R., Clark S.C.;
RL Submitted (SEP-1996) to the EMBL/Genbank/DBJ databases.

DR EMBL, U70136; AAB09089.1; -.
DR Genew; HGNC:9364; PRG4.
DR InterPro; IPR000585; Hemopexin.
DR InterPro; IPR002965; P_rich_extensn.
DR InterPro; IPR001212; Somatomedin_B.
DR Pfam; PF00045; hemopexin; 2.
DR Pfam; PF01033; Somatomedin_B; 2.
DR PRINTS; PR01217; PRICHEXTENSIN.
DR PRINTS; PR00022; SOMATOMEDINB.
DR SMART; SMO0120; HK; 2.
DR SMART; SMO0201; SO; 2.
DR PROSITE; PS00024; HEMOPEXIN; 1.
DR PROSITE; PS00524; SOMATOMEDIN_B; 2.
SQ SEQUENCE 1404 AA; 151090 MW; AABD7AD19B35F4F6 CXC64

Query Match	100.0%;	Score 7526;	DB 4;	Length 1404;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 1404;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

Qy	1	NAMKTLPIYLILLLLSFVLIQOVSODLS	SSCAGCGGYSNDAICNDYNOQHMECCPDF	60
Dp	1	NAMKTLPIYLILLLLSFVLIQOVSODLS	SSCAGCGGYSNDAICNDYNOQHMECCPDF	60
Qy	61	KRVCTAEIASCRCESFERGECDDAOCKYDKCCPDYEFSCAEVHNPTSPSSKKAP	120	
Dp	61	KRVCTAEIASCRCESFERGECDDAOCKYDKCCPDYEFSCAEVHNPTSPSSKKAP	120	
Qy	121	PPSGASQITIKSTTKCS	SPKPNKKTKYIESEETIEHSVSENQESSSSSSSSSSSTIW	180
Dp	121	PPSGASQITIKSTTKCS	SPKPNKKTKYIESEETIEHSVSENQESSSSSSSSSSSTIW	180
Qy	181	KIKSSKNSAANBELQKULKVKONKKNRTKKK	PPKRPVYDEASSGJDNDFKTYTDPST	240
Dp	181	KIKSSKNSAANBELQKULKVKONKKNRTKKK	PPKRPVYDEASSGJDNDFKTYTDPST	240
Qy	241	TOHNKYSTPKITTAAPINPRPSL	PNSDPTSKEITSLTVNKETVEHKEITTTNKQSTDG	300
Dp	241	TOHNKYSTPKITTAAPINPRPSL	PNSDPTSKEITSLTVNKETVEHKEITTTNKQSTDG	300
Qy	301	KEKITSABETOSIEKITSANDLAPTS	KVLAKPTPKAETTTKGPAJTTPKKEPTTPKEBPAS	360
Dp	301	KEKITSABETOSIEKITSANDLAPTS	KVLAKPTPKAETTTKGPAJTTPKKEPTTPKEBPAS	360
Qy	361	TTPKKEPTTPTKSAP	PTTPKEBPATTTKSAPTTPKEBPATTTKEBPATTTKEBPATTTKEP	420
Dp	361	TTPKKEPTTPTKSAP	PTTPKEBPATTTKSAPTTPKEBPATTTKEBPATTTKEBPATTTKEP	420
Qy	421	APTTKSAP	TPPKKEBPATTPPKKAPPTTPKEBPATTPKEBPATTPKEBPATTPKEBPATTPK	480
Dp	421	APTTKSAP	TPPKKEBPATTPPKKAPPTTPKEBPATTPKEBPATTPKEBPATTPKEBPATTPK	480
Qy	481	EPAPAPAKKAPATTPKEBPATTPKEBPATTTKEBSP	PTTPKEBPATTTKSAPTTTKEBPAPT	540
Dp	481	EPAPAPAKKAPATTPKEBPATTPKEBPATTTKEBSP	PTTPKEBPATTTKSAPTTTKEBPAPT	540
Qy	541	TTKSAPTTPKEBSP	PTTPKEBPATTPKEBPATTPPKKAPPTTPKEBPATTPKEBPATTTKKP	600
Dp	541	TTKSAPTTPKEBSP	PTTPKEBPATTPKEBPATTPPKKAPPTTPKEBPATTPKEBPATTTKKP	600
Qy	601	APTAPKEBPATTPKETA	PTTPKULPTTPBKLAPTTPKEBPATTPBEELAPTTPBEETPTT	660
Dp	601	APTAPKEBPATTPKETA	PTTPKULPTTPBKLAPTTPKEBPATTPBEELAPTTPBEETPTT	660
Qy	661	PEBPAPTPPKAAAPMTPKEBPATTPKEBPATTPKEBPATTPKETA	PTTPKGAPTTLKEP	720
Dp	661	PEBPAPTPPKAAAPMTPKEBPATTPKEBPATTPKEBPATTPKETA	PTTPKGAPTTLKEP	720
Qy	721	APTTPPKAPKELAPTTTKEBPTSTSDKAP	PTTPKGTAPTTPKGAPTTPKG	780
Dp	721	APTTPPKAPKELAPTTTKEBPTSTSDKAP	PTTPKGTAPTTPKGAPTTPKG	780
Qy	781	TAPTTLKEBPATTPKKAPAKELAPTTTKGP	STSTSDKAPPTTPKETAPTTPKEBPATTPK	840

Db	761	TAAPTLKBPAPPTPPKAPKAPKELAPPTTKGPTSTSDKAPPTTPKETAAPPTTKBPAPPTTK	840
Qy	841	KPAPTPEPPTTSEVSTPTTTKEPTTIHKSPDSETPELSAEPFKALENSPKEGVPT	900
Db	841	KPAPTPEPPTTSEVSTPTTTKEPTTIHKSPDSETPELSAEPFKALENSPKEGVPT	900
Qy	901	TKTPAATPEMNTTAKOKTTTEDLATTPBTTTAAAPMTKETAATTTKTESKLTATTTQV	960
Db	901	TKTPAATPEMNTTAKDKTTERDLATTPBTTTAAAPMTKETAATTTKTESKLTATTTQV	960
Qy	961	TSTTQDPTPFKITLTKTTTLAPKTYTTTKITITTEINMKPEETAAPKORATNSKATTPK	102
Db	961	TSTTQDPTPFKITLTKTTTLAPKTYTTTKITITTEINMKPEETAAPKORATNSKATTPK	102
Qy	1021	POKPTPAKPKPTSTYKPKTMPBRVKRKPPTTPRKMSTMPBELNPTSRIAEAMLQTTTRPN	108
Db	1021	POKPTPAKPKPTSTYKPKTMPBRVKRKPPTTPRKMSTMPBELNPTSRIAEAMLQTTTRPN	108
Qy	1081	QTPNKLVEVNPKSESDAGABEETPHMLRPVFMPEVTPDMDYLPRVNOGIINPMLS	114
Db	1081	QTPNKLVEVNPKSESDAGABEETPHMLRPVFMPEVTPDMDYLPRVNOGIINPMLS	114
Qy	1141	DETNI.CNGSPVDGLTTLNRGTIVARFGHYFMMLSFSPSPSPARRITEVNGISPIDTVEF	120
Db	1141	DETNI.CNGSPVDGLTTLNRGTIVARFGHYFMMLSFSPSPSPARRITEVNGISPIDTVEF	120
Qy	1201	RCNCGKGTFFFKDSQYMRFTNDIKDAGYKPIFKGFGGLTGQIIVALSTAKYKNMDESY	126
Db	1201	RCNCGKGTFFFKDSQYMRFTNDIKDAGYKPIFKGFGGLTGQIIVALSTAKYKNMDESY	126
Qy	1261	FFKRGGSIQOYTYKQEPYQKCGRRPALNYPYEGMTQVRRRRFERAIGPSQHTTIRIQY	132
Db	1261	FFKRGGSIQOYTYKQEPYQKCGRRPALNYPYEGMTQVRRRRFERAIGPSQHTTIRIQY	132
Qy	1321	SPARLAYODKGLVHNEVVSILMRGLPNVVSATSLPNIRKDGXYDYAFSKDOQYNIQV	138
Db	1321	SPARLAYODKGLVHNEVVSILMRGLPNVVSATSLPNIRKDGXYDYAFSKDOQYNIQV	138
Qy	1381	PSRTARAITTSSGQTLISKVWVNC	1404
Db	1381	PSRTARAITTSSGQTLISKVWVNC	1404
RESULT 2			
Q9BX49			
AC	Q9BX49	PRELIMINARY:	PRT: 1404 AA.
DT	01-JUN-2001	(TREMblrel. 17, Created)	
DT	01-JUN-2001	(TREMblrel. 17, Last sequence update)	
DT	01-MAR-2003	(TREMblrel. 23, Last annotation update)	
DE	BGI7416.2	(MSF: megakaryocyte stimulating factor).	
GN	BGI7416.2.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_Taxid=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	Wray P.;		
RL	Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AL133553; CAC36090.1; -		
DR	InterPro; IPR000585; Hemopexin.		
DR	InterPro; IPR002965; P rich extenxn.		
DR	InterPro; IPR001212; Somatomedin_B.		
DR	Pfam; PF00045; hemopexin; 2.		
DR	Pfam; PF01033; Somatomedin_B; 2.		
DR	PRINTS; PR01217; PRICHEXTENS.		
DR	PRINTS; PR00022; SOMATOMEDINB.		
DR	SMART; SM00201; SO; 2.		
DR	SMART; SM00201; SO; 2.		
DR	PROSITE; PS00024; HEMOPEXIN; 1.		
DR	PROSITE; PS00524; SOMATOMEDIN_B; 2.		
SO	SEQUENCE	1404 AA; 151076 MW; 782A11746A3FDBE5 CRC64;	

Query Match 99.6%; Score 7497; DB 4; Length 1404;
 Best Local Similarity 99.7%; Pred. No. 0;
 Matches 1400; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

QY 1 MAMKTLPIYLLLLSVFIQOVSSODLSSCAGRCGEGSRDATCNCDCYNCOHYMECCPDF 60
DB 1 MAMKTLPIYLLLLSVFIQOVSSODLSSCAGRCGEGSRDATCNCDCYNCOHYMECCPDF 60
QY 61 KAVCTAELSCGRCFESFERGRCDDAQCCKTDKCCPDYESTCAVYANPTSPSSKAP 120
DB 61 KAVCTAELSCGRCFESFERGRCDDAQCCKTDKCCPDYESTCAVYANPTSPSSKAP 120
QY 121 PPSGASQTIKSTTKRSPKPNKKTKKVISEEITTEHSVSEKQESSSSSSSSSSSTIM 180
DB 121 PPSGASQTIKSTTKRSPKPNKKTKKVISEEITTEHSVSEKQESSSSSSSSSSSTIR 180
QY 181 KIKSSKNSAANRELQKLVKDKNKKRTRKKKPTPKPVVDAGSGLDNGDFKVTPTDST 240
DB 181 KIKSSKNSAANRELQKLVKDKNKKRTRKKKPTPKPVVDAGSGLDNGDFKVTPTDST 240
QY 241 TQHNKISTSPKITTAAPINRPSLPNSDTSKETSITVNKETVETKETTTNKQTSIDG 300
DB 241 TQHNKISTSPKITTAAPINRPSLPNSDTSKETSITVNKETVETKETTTNKQTSIDG 300
QY 301 KEKTTSAKETSIEKTSANDLAFTSKVLAKPPKATTTTGGALTTPKPTTTPKEPAS 360
DB 301 KEKTTSAKETSIEKTSANDLAFTSKVLAKPPKATTTTGGALTTPKPTTTPKEPAS 360
QY 361 TTPKEPTPTTIKSAPTTPKEBAPTITKSAPTTPKEBAPTITTPKEBAPTITTPKE 420
DB 361 TTPKEPTPTTIKSAPTTPKEBAPTITKSAPTTPKEBAPTITTPKEBAPTITTPKE 420
QY 421 APTTTSATTPKEBAPTTPPKAPATTTPKEBAPTTPKEBAPTTPKEBAPTTPK 480
DB 421 APTTTSATTPKEBAPTTPPKAPATTTPKEBAPTTPKEBAPTTPKEBAPTTPK 480
QY 481 BPAPTAPKAPATTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPT 540
DB 481 BPAPTAPKAPATTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPT 540
QY 541 TTKSATTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPT 600
DB 541 TTKSATTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPT 600
QY 601 APTAPKEBAPTTPKEBAPTTPPKLJPTTPEKLAPTTPKEBAPTTPPEEPTPT 660
DB 601 APTAPKEBAPTTPKEBAPTTPPKLJPTTPEKLAPTTPKEBAPTTPPEEPTPT 660
QY 661 PSEBAPTTPKAAAPNTPKBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPT 720
DB 661 PSEBAPTTPKAAAPNTPKBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPT 720
QY 721 APTTPKAPKELAPTTKEPTSTSDKAPATTPKGTATTPKEBAPTTPKEBAPTTPK 780
DB 721 APTTPKAPKELAPTTKEPTSTSDKAPATTPKGTATTPKEBAPTTPKEBAPTTPK 780
QY 781 TAPTLKEBAPTTPKAPKELAPTTKGPSTSTSDKAPATTPKGTATTPKEBAPTTPK 840
DB 781 TAPTLKEBAPTTPKAPKELAPTTKGPSTSTSDKAPATTPKGTATTPKEBAPTTPK 840
QY 841 KAPAPTTPPTTSSVSTPTTKEPTTHKSDESTPELSAPPTKALENSPEKEGVPT 900
DB 841 KAPAPTTPPTTSSVSTPTTKEPTTHKSDESTPELSAPPTKALENSPEKEGVPT 900
QY 901 TKTPTATKEMTTTADKTERDLRTPTETTTAAPTKEIATTTTEKTESKITATTVOY 960
DB 901 TKTPTATKEMTTTADKTERDLRTPTETTTAAPTKEIATTTTEKTESKITATTVOY 960
QY 961 TSTTTODTTPPKITTLKTTTLAPKVTYTKYITTEIANKPEETAKKORATNSKATTPK 1020
DB 961 TSTTTODTTPPKITTLKTTTLAPKVTYTKYITTEIANKPEETAKKORATNSKATTPK 1020

```

```

QY 1021 POKPTKAPKPKSTTKPKTPMPVRKPKTTPTRKQTSIMPELNPSTRIAEALQTTTRN 1080
DB 1021 POKPTKAPKPKSTTKPKTPMPVRKPKTTPTRKQTSIMPELNPSTRIAEALQTTTRN 1080
QY 1081 QTPNSKLVENVKESDAGAEGETPHMLLRPHVEMPEVTDMVDYLRVNOGIIINPMLS 1140
DB 1081 QTPNSKLVENVKESDAGAEGETPHMLLRPHVEMPEVTDMVDYLRVNOGIIINPMLS 1140
QY 1141 DETNTCNKRPVUGLTLTNGTLVAFRGHYFMWLSFSPSPSPARRITTEWGISPTDVT 1200
DB 1141 DETNTCNKRPVUGLTLTNGTLVAFRGHYFMWLSFSPSPSPARRITTEWGISPTDVT 1200
QY 1201 RCNCEGKTFEFDQSYMFRTNDIKDAGYKPIFKGFGGLTGOIYVALSTAKKMPESY 1260
DB 1201 RCNCEGKTFEFDQSYMFRTNDIKDAGYKPIFKGFGGLTGOIYVALSTAKKMPESY 1260
QY 1261 PFKRGSIQOYLYKQEPVQKCGRRPALNYPVYGEMTOVRRRRFERAIGPSQTHIRIQ 1320
DB 1261 PFKRGSIQOYLYKQEPVQKCGRRPALNYPVYGEMTOVRRRRFERAIGPSQTHIRIQ 1320
QY 1321 SPARLAYODKVLHNEVKVSIIMRGLPNVYTSALSIPNIRKPDGYDYAFSKOYYNIDV 1380
DB 1321 SPARLAYODKVLHNEVKVSIIMRGLPNVYTSALSIPNIRKPDGYDYAFSKOYYNIDV 1380
QY 1381 PSRTARATTRSGOTLSKWNCP 1404
DB 1381 PSRTARATTRSGOTLSKWNCP 1404

```

```

RESULT 3
Q9JW99 PRELIMINARY; PRT: 1054 AA.
ID Q9JW99
AC Q9JW99;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypochemical 116.0 kDa protein.
GN PRG4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20573856; PubMed=11124536;
RA Ikegawa S., Sano M., Koshizuka Y., Nakamura Y.;
RT "Isolation, characterization and mapping of the mouse and human PRG4
RT (proteoglycan 4) genes.";
RL Cytogenet. Cell Genet. 90:291-297(2000).
DR EMBL: AB034730; BAA92310.1; -.
DR MGD: MGI:1891344; Prg4.
DR InterPro: IPR000585; Hemopexin.
DR InterPro: IPR002965; P rich exten.
DR InterPro: IPR001212; Somatomedin_B.
DR Pfam: PF00045; hemopexin_2.
DR Pfam: PF01033; Somatomedin_B_2.
DR PRINTS: PRO1217; PRICHEXTNSN.
DR PRINTS: PRO0022; SOMATOMEDINB.
DR SMART: SM00120; HX; 2.
DR SMART: SM00201; SO; 2.
DR PROSITE: PS00024; HEMOPEXIN_1.
DR PROSITE: PS00524; SOMATOMEDIN_B_2.
SQ SEQUENCE 1054 AA; 11592 MW; 4FC64BFA42283235 CRC64;

```

Query Match 53.0%; Score 3988; DB 11; Length 1054;
 Best Local Similarity 57.3%; Pred. No. 1.3e-209;
 Matches 811; Conservative 59; Mismatches 172; Indels 374; Gaps 25;

```

QY 1 MAMKTLPIYLLLLSVFIQOVSSODLSSCAGRCGEGSRDATCNCDCYNCOHYMECCPDF 60
DB 1 MAMKTLPIYLLLLSVFIQOVSSODLSSCAGRCGEGSRDATCNCDCYNCOHYMECCPDF 60
QY 61 KAVCTAELSCGRCFESFERGRCDDAQCCKYDKCCPDYESTCAVYANPTSPSSKAP 120

```



```

Db      61 KRVCSPBELSCGRGCFESFARGRECDSDQCKQYGGCCADYDSFCHEVANSUS-PSKXAP 119
Qy      121 PPSGASQITKSTTKRSPKPKKTKTKYIBSEBEITEHSVSENQSSSSSSSSSSSTW 180
Db      120 TPAGASDITKSTTKRSPKSPKT-TRITKYVESEBELTEHSDSENOG--SSSSSSSSSSSTIR 176
Qy      181 KIKSSKNSANRELOKLVKONKNKRTKKKPTPKPVVDEAGSGLDJGDGDFKVT--TPDT 238
Db      177 KIKSSKNS-ANRELOKNSVNDKNNKTKPKKPNPEPPVDEAGSGLDJGDGDFKLTPEPPDP 235
Qy      239 STTOHKNVSTSPKITTAAPINRPSLPNSDTSKESLTVNKETVETKETTNNKQST 298
Db      236 PTTHSKAVTSPKTTAAKPVTPKPSLAPNSETSKEASLASNKETTVETKETTATNKQSSA 295
Qy      299 DGEKTSKAKTOSIEKTSADOLAFSTKVLAKPPTKAEITTKGPAITPKPEPTTPKEP 358
Db      296 -SKKKTTSVKETRSKETSDDOV----- 319
Qy      359 ASTPKPEPTPTTKSAPTPKKPAPTTKSAPTPKKPAPTTKKPAPTTKKPAPTTK 418
Db      320 TSTTPK-----NSAPTTTKPV-TTKSKKFLD-----LQPEPTTK 357
Qy      419 EPAPTTKSAPTTKKPAPTTKKPAPTTKKPAPTTKKPEPTTKKPAPTTKKPAPTT 478
Db      358 EPPPTTKKPEPTTKKPEPTTKKPEPTTKKPEPTTKKPEPTTKKPEPTTKKPEPTT 417
Qy      479 PKKPAPTPKKPAPTTKKPAPTTKKPAPTTKKPESTTKKPAPTTKKSAPTTKKPA 538
Db      418 PKKPEPTPKKPEPTTKKPEPTTKKPESTTKKPEPTTKKPEPTTKKPEPTTKKPEPT 461
Qy      539 PTTTKSAPTTKKPEPTTKKPAPTTKKPAPTTKKPAPTTKKPAPTTKKPAPTTK 598
Db      462 -----EPTTKKPEPTTKKPEPTTKKPEPTTKKPEPTTKKPEPTTKKPEPTT 512
Qy      599 KPAPAPKPAPTTKKPAPTTKKPAPTTKKPAPTTKKPAPTTKKPAPTTKKPEPTP 658
Db      513 ----- 512
Qy      659 TTPKPAPTTKKAAAPTKKPAPTTKKPAPTTKKPAPTTKKPAPTTKKPAPTTK 718
Db      513 -----TPKPEPTTKKPEPTTKKPEPTTKKPEPTTKKPEPTTKKPEPTTK 539
Qy      719 EPAPTTPKKPAKELAPTTKBPSTSDKAPPTPKTAAPTTKPAPTTKKPAPTTK 778
Db      540 EPPPTTKPKP-----EPTTKKPEPTTKKPEPTTKKPEPTTKKPEPTTK 570
Qy      779 KGTAPTTKKEAPTTPKKPAKELAPTTTKGPTSTSDKPAPTTKKPAPTTKKPAPTT 838
Db      571 -----KEPEPTTKKPEPTTKKPEPTTKKPEPTTKKPEPTTKKPEPTTK 609
Qy      839 PKKPAPTPKKPEPTTKKPEPTTKKPEPTTKKPEPTTKKPEPTTKKPEPTTK 898
Db      610 PKKPEPTTK----- 617
Qy      899 PTTKPAATKPEMTTAKDKTBRDLRTTPETTTAAPKMTKELATTEKTESKITATTT 958
Db      618 -----TSPKTI----- 622
Qy      959 QVSTTTODTTPFKITTLTKTTTLAKVTTTKKITTTTEINMKPESTAKPKDRAATNSKAT 1017
Db      623 -----TLKATTLAKVTAPAE-----EIQKPEPTTKPASESDSDSKITL 662
Qy      1018 -----TPKP-QKTKAPKPKPTSTKKPKTMKPVKPKPTTTPPKMSTMBELNFTSIA 1069
Db      663 KPQKPTKAPKPKPTSTKKPKPTSTKKPKT-PTKTKPKTTTAPALKTISATBELNFTT-P 719
Qy      1070 EAMQTTTRPMQTPSKLVEVNPKSSEDAGABGEPHMLLRPHVMEPVTDMDVLRVP 1129
Db      720 EVMPLPTTIPKQTPPEPTAENVPDHEDADGGEKAP-LIPGPVLPFPALBGTDLAQL 778
Qy      1130 NOGIINMLSDETNINCCKEVDGLTTLRNGTLVFRGHYFWMLSFPSPSPARRITEVM 1189

```

```

Db      779 NRGINIMPSEDETNLGCKEVDGLTTLRNGTLVFRGHYFWMILNPRPSPARRITEVM 838
Qy      1190 GISPIDTVPTRCNCEKTEPFKDSOYMRFTNDIKDGYRPIPKFGGLTGOI VALST 1249
Db      839 GISPIDTVPTRCNCEKTEPFKDSOYMRFTNDVDGYPQIVKGFGGLTGI VALST 898
Qy      1250 AKYNMPESSVYFFRGSGSIQYIYKQEPVQKCPGRPALNYPVYGMTOVRRRFEPAIG 1309
Db      899 AKYNDPESVYFFRGSGSIQYIYKQEPVQKCPGRPALNYPVYGMTOVRRRFEPAIG 958
Qy      1310 PSQHTTRIOYS-PALAYQDKVLANEVKYSILMRGLPNVTIAISLPNTRKPDGDY 1368
Db      959 PFQHTTRIHYSVPMRVSYODKGLHNEVKYSTMMRGFPNVVTSAILPNTRKPDGDY 1018
Qy      1369 AFSKQYVNIIDVPSRTARITTRSGOTLSKVWVNC 1404
Db      1019 AFSKQYVNIIDVPSRTARITTRSGOTLSKIWNCP 1054

RESULT 4
ID 07765 PRELIMINARY; PRT; 401 AA.
AC 07765;
DT 01-NOV-1998 (TRENBLrel. 08, Created)
DT 01-NOV-1998 (TRENBLrel. 08, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE Superficial zone protein (Fragment).
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
NC NCHI_Taxid=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Articular cartilage;
RX MEDLINE=99171663; PubMed=920774;
RA Flannery C.R., Hughes C.E., Schumacher B.L., Tudor D., Aydelotte M.B.,
RA Kuetner K.E., Caterson B.;
RT "Articular cartilage superficial zone protein (SZP) is homologous to
RT megakaryocyte stimulating factor precursor and is a multifunctional
RT lubricating protein with potential growth-promoting, cytoprotective, and
RT biochemical properties in cartilage metabolism."
RL Biochem. Biophys. Res. Commun. 254:535-541 (1999).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Articular cartilage;
RX MEDLINE=99171663; PubMed=10073655;
RA Schumacher B.L., Hughes C.E., Kuetner K.E., Caterson B.,
RA Aydelotte M.B.;
RT "Immunodetection and partial cDNA sequence of the proteoglycan,
RT superficial zone protein, synthesized by cells lining synovial
RT joints."
RL J. Orthop. Res. 17:110-120 (1999).
DR EMBL; AF056218; AAD13404.1; -.
DR InterPro; IPR00585; Hemoexin.
DR InterPro; IPR002965; P-rich_extensn.
DR Pfam; PF00045; hemoexin; 2.
DR PRINTS; PR01217; PRICHEXTENS.
DR SMART; SM00120; HX; 2.
DR PROSITE; PS00024; HEMOEXIN; 1.
FT NON TER 1
SQ SEQUENCE 401 AA; 44952 MW; 86147CC9AFB73D7 CRC64;

Query Match 22.8%; Score 1713.5; DB 6; Length 401;
Best Local Similarity 80.2%; Pred. No. 5e-86;
Matches 320; Conservative 24; Mismatches 52; Indels 3; Gaps 3;

Qy      1007 PKDASTSKATTPKQKPTAKPKKPTSTKKPKTMKPVKPKPTTTPPK-MSTMBELNPT 1065
Db      5 PKGATNSQVTPKQKPTAKPKKPTSTKKPT-PRVRKPKTTTTPPTTTSANMBEPTPT 63
Qy      1066 SRIEAMQTTTRNQTPNSKLVEVNPKSSEDAGABGEPHMLLRPHVMEPVTDMDYL 1125

```


Db 305 T-TTTRCTTTTSTCTRRPTTTTPTSTTTTCTSGTTTPTSTTTTCTSGTTTPTSTTTT 363
Qy 430 TTPKAP-PAPTPPK-----PAPTPKAPPTTPKAPPTTPKAPPT-TKAPAPTP 479
Db 364 TTTSGPTTTTPTSTTTTSTSGPTTTTPTSTTTTSTSGPTTTTPTSTTTTSTSGPTTTT 423
Qy 480 KEPAPTPKAPPTTPKE-----PAPTPKAPPTTPKAPPTTPKAPPTTPKAP 531
Db 424 RSTTTTSGPTTTTPTSTTTTSTSGPTTTTPTSTTTTSTSGPTTTTPTSTTTTSTSGPT 483
Qy 532 TTPKAPPTTPKAP--TTPKAPPTTPKAPPTTPKAP-----PAPTPKAPPTTP 581
Db 484 TTTPTSTTTTSTSGPTTTTPTSTTTTSTSGPTTTTPTSTTTTSTSGPTTTTPTSTTTTST 543
Qy 582 KEPAPTPKAPPTTPKAPAPK-----PAPTPKAPPTTPKAPPTTPKAP 633
Db 544 SGPTTTTPTSTTTTSTSGPTTTTPTSTTTTSTSGPTTTTPTSTTTTSTSGPTTTTPTST 603
Qy 634 PTPKAPPTTPKAPPTTPKAPPTTPKAP-----PAPTPKAPPTTPKAPPTTP 685
Db 604 TTTSGPTTTTPTSTTTTSTSGPTTTTPTSTTTTSTSGPTTTTPTSTTTTSTSGPTTTTPT 663
Qy 686 KEPAPTPKAPPTTPKET-----APTPKAPPTTPKAPPTTPKAPPTTPKAP 737
Db 664 RSTTTTSGPTTTTPTSTTTTSTSGPTTTTPTSTTTTSTSGPTTTTPTSTTTTSTSGPT 723
Qy 738 TKEPTSTSDK---PAPTPKAPPTTPKAPPTTPKAP-----PAPTPKAPPTTP 786
Db 724 TTTPTSTTTTSTSGPTTTTPTSTTTTSTSGPTTTTPTSTTTTSTSGPTTTTPTSTTTTST 783
Qy 787 KEPAPTPPK-----PAPKAPPTTPKAPPTTP-----SDKAPPTTPKETA 827
Db 784 SGPTTTTPTSTTTTSTSGPTTTTPTSTTTTSTSGPTTTTPTSTTTTSTSGPTTTTPTST 843
Qy 828 PTPKAPPTTPKAPPTTPKAPPTTPKAPPTTPKAPPTTPKAPPTTPKAPPTTPKAPPTTP 887
Db 844 TTTSGPTTTTPTSTTTTCTSCPTTTTPTSTTTTCTSCPTT--TTPRSTTTTCTSGPTTT 901
Qy 888 ALENSPK-EPGVPTTTPKAPKAPPTTPKAPPTTPKAPPTTPKAPPTTPKAPPTTPKAP 946
Db 902 TPTSTTTTCTSCAPPTTPRST---TTPSTTPPTTPPTSTTTTPTSTTTTPTSTTTPT 957
Qy 947 KTSKSKATTTTCTTODTTPPTTPKAPPTTPKAPPTTPKAPPTTPKAPPTTPKAPPTTPKAP 1005
Db 958 STSGPTTTTPTSTTTTCTSCPTTTTPTSTTTTCTSCAPPTTPPTSTTTT--TSRPTTTP 1015
Qy 1006 KPKORATNSKATTPKAPK--PTKAPKAPPTTPKAPPTTPKAPPTTPKAPPTTPKAPPTTP 1064
Db 1016 RSTTTTPTSTTTTPTSTTTPTCTSRPTTTPTSTTTTPTSTTTTPTSTTTTPTSTTTPT 1075
Qy 1065 TSRIAEMLQTTT-----PQOT 1083
Db 1076 S---ASPTTPTTTPPC 1100

RESULT 6
Q9N4S7 PRELIMINARY; PRT, 1079 AA.
AC Q9N4S7;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Y5IB11A.1 protein.
GN Y5IB11A.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderiinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RC SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=99069613; PubMed=9851916;

RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Cotton M.1
RT "The sequence of C. elegans cosmid Y5IB11A.";
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Waterston R.1
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC006797; AF60743.1;
DR Wormpep; Y5IB11A.1; CE22326.
DR InterPro; IPR002965; P. rich. extensn.
DR PRINTS; PR01217; PRICHEXTNSN.
SQ SEQUENCE 1079 AA; 110532 MW; 8DBDE3824CF80CA1 CRC64;

Query Match. 13.1%; Score 982.5; DB:5; Length 1079;
Best Local Similarity 29.6%; Pred. No. 1.1e-45;
Matches 332; Conservative 131; Mismatches 481; Indels 177; Gaps 44;

Qy 195 QKLVKNDKNNKRRKPKPPVVDAGSLDNGDFV-----TTPDSTTQ-HNKV 246
Db 27 QKLELDICRANQPPHMLP-----STLSVDNETBSTVLVSTPSSSTPIKETT 77
Qy 247 STSPKITTAKKINRPSLPNSDTSKETSIVNKEET-VETKETTNNKQSTDGKETT 305
Db 78 TTAETTTSTP-----PSSSTTPVQTTTAAETTTSTAPSSSTTPVQTT-----TT 124
Qy 306 SAKTOSIEKTSADLAFTSVLAKPDKAETT-TKAP-ALTTKEPPTTPKAPASTTP 363
Db 125 TAPETTTSTPSSS-----TSPVQTTTAAETTTSTAPSSSTTPVQTTTAAETTTSTEP 180
Qy 364 KEPPTTPKAPPTTPKAPPTTPKAPPTTPKAPPTTPKAPPTTPKAPPTTPKAPPTTPKAP 423
Db 181 PSSSTTPVQTTTAAETTTSTPSSSTTPVQTTTAAETTTSTP-----TTSSTP-PSSSTTPVQ 236
Qy 424 TTKAPPTTPKAPPTTPKAPPTTPKAPPTTPKAPPTTPKAPPTTPKAPPTTPKAPPTTPKAP 483
Db 237 TTTTAAETTTSTPSSSTTPVQTTTAAETTTSTEP-PSSSTTPVQTTTAAETTTSTEP 295
Qy 484 PTAKKAPPTTPKAPPTTPKAP-----APTTKAPPTTPKAPPTTPKAPPTTPKAPPTTPKAP 529
Db 296 PSSSTTPVQTTTAAETTTSTPSSSTTPVQTTTAAETTTSTPSSSTTPVQTTT 355
Qy 530 APPTTK--EPAPTTTKAPPTTPKAPPTTPKAPPTTPKAPPTTPKAPPTTPKAPPTTPKAP 586
Db 356 APETTTREPSSSTTPVQTTTAAETTTSTPSSSTTPVQTTT-----APETTTSTEP 411
Qy 587 TTPKAPPTTPKAPPTTPKAPKAP--APTPKETAPTTPKAPPTTPKAPPTTPKAPPTTPKAP 644
Db 412 SSSSTTPVQTTTAAETTTSTPSSSTTPVQTTTAAETTTSTEP-----PSSSTTPVQTT 467
Qy 645 PEAAPTTPPEPTTTPKAPPTTPKAPAAAPNTKAP--APTPKAPPTTPKAPPTTPKAP 702
Db 468 TTPAETTTSTEP-PSSSTTPVQTTTAAETTTSTPSSSTTPVQTTTAAETTTSTEP-TTSTES 525
Qy 703 ETAPTTKGAAPTTLKAPPTTPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAP 761
Db 526 PSSSTTPVQTTTAAETTTSTPSSSTTPVQTTTAAETTTSTPSSSTTPVQTTT 585
Qy 762 PKKAPPTTPKAP--APTPKGAAPTTLKAPPTTPKAPKAPKAPKAPKAPKAPKAPKAPKAP 818
Db 586 APK-----TTSSTREPSSSTTPVQTTTAAETTTSTPSSSTTPVQTTTAAETTTSTEP 642
Qy 819 APPTPKETAAPTTPKAPPTTPKAP-----PAPTPETPTPTTS-----EVSPTTTKE 865
Db 643 SSTTPVQTTTAAETTTSTEPSSSTTPVQTTTAAETTTSTEPSSSTTPVQTTTAAETTTTAP 702


```

QY 371 IKSAPTTKEBA---PTTKSAPTTKEBAPT---PTTKBAPT---PTTKBAPT 424
Db 2523 LKPTBGTAKPTTLKPTBGTSAKPTTLKPTBGTSAKPTTLKPTBGT 2582
QY 425 TKSAPTTKEBAPTTPK---KPAPTTKEBAPTTPKSP---PTTKBAPTKEBAPT 477
Db 2583 AK--PTTLKPTBGTAKPTTLKPTBGTSAKPTTLKPTBGTAKPTTLKPTBGTSAKPTTL 2640
QY 478 TPKEBAPTAPKPAPTTKEBAPTTPKEBAPTTPKEB---PTTKBAPTTPKSAPTT 534
Db 2641 KPTBGTAKPTTLKPTBGTSAKPTTLKPTBGTAKPTTLKPTBGTSAKPTTLKPTBGT 2699
QY 535 KEBA---PTTKSAPTTKEBAPTTPK---PTTKBAPTTPK---KP 576
Db 2700 AKPTTLKPTBGTSAKPTTLKPTBGTAKPTTLKPTBGTAKPTTLKPTBGTAKPTTLK 2759
QY 577 A-----PTTKBAPTTPK---EPAPTTPKPA---PTAKBAPTTPKETAPTT--P 621
Db 2760 TKGTSAPKPTTLKPTBGTAKPTTLKPTBGTAKPTTLKPTBGTSAKPTTLKPTBGTAKP 2819
QY 622 KKLPT---TPKLAAPTTPKEBAPT---PEBLAPT--TPBPTPTTPREP 664
Db 2820 TTLKPTBGTAKPTTLKPTBGTSAKPTTLKPTBGTAKPTTLKPTBGTSAKPTTLKPTBGT 2879
QY 665 APPTPKAAAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPT 714
Db 2880 TTAKPTTLKPTBGTSAKPTTLKPTBGTAK---PTTLKPTBGTAKPTTLKPTBGTSAK 2936
QY 715 TTLKBPAPTTPK---KEBAPBAPTTPKPTSTSDKPA-----PTTKBGTAP 759
Db 2937 TTLKPTBGTAKPTTLKPTBGTAKPTTLKPTBGTSAKPTTLKPTBGTAKPTTLKPTBGT 2996
QY 760 TPKEBAPTTPKEBAPTTPKGTAPTTLKEBAPTTPK---KPAKELAPTTPKPTSTTS 815
Db 2997 TTAK---PTTLK---PTBGTSAKPTTLKPTBGTAKPTTLKPTBGTSAKPTTLKPTBGT 3050
QY 816 DKPA-----PTTKBAPTTPKEBAPT---PKKAPAPTTPPTPTSTSTPT 862
Db 3051 AKPTTLKPTBGTSAKPTTLKPTBGTAK---PTTLKPTBGTSAKPTTLKPTBGTAKPTT 3107
QY 863 TK-----BPTTIKSPDSTPELSAPTTP--KALENSPKBPGV-----PTTKBAPTAPK 909
Db 3108 LKPTBGTSAKPTTL--KPTBGT---TAKPTTLKPTBGTAKPTTLKPTBGTAKPTTLK 3162
QY 910 EMTTAKDKTERDLKPTTPTT-----AAPTKEKATTTTEKTESKIT--ATTTOV 960
Db 3163 TEGTSAPKPTTLKPTBGTAKPTTLKPTBGTSAKPTTLKPTBGTAKPTTLKPTBGTSAOP 3222
QY 961 TSTTODTTPKITTTL-----KTTTLAPKVTTPK--TTTTTEIMNKKEBTAAPKDR 1010
Db 3223 TTLKPTBGTAKPTTLKPTBGTSAKPTTLKPTBGTAKPTTLKPTBGTSAOPPTTLKPTD 3281
QY 1011 ATNSKATTPKPKOKPTKAPK---KPT--STKPKTMP---RVKPKPTTPRKMSTWMP 1060
Db 3282 GTAKPTTLKPTBGTSAKPTTLKPTBGTAKPTTLKPTBGTAKPTTLKPTBGTSAKPT 3340
QY 1061 ELAPTSAEAMLOTTTRPNOTPNKSKAVEVNPKSDAGAGETPPHMLRPHVPMPEVTP 1120
Db 3341 TLKPTL--GTAKPTTLKPTBGTSAKPTTLK---TEGTT---AKPTTLKPTBGT 3387
QY 1121 DMDYLPVNPNGIILNPMLSDETMINCNKGPVDSLTTLR 1158
Db 3388 SAKPTTLKPTBGTAKP---TTL---KPTDGTTPPR 3417

```

```

DE CG32602-PB.
GN CG32602.
OC Drosophila melanogaster (fruit fly).
OC Euxariota, Metazoa, Arthropoda, Hexapoda, Insecta, Pezomyzeta;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoeklin R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H., Blazer V.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
RA Abell J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballwey R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
RA Butrie K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferreira C., Ferreira S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hosteln D., Houston K.A., Howland T.J., Wei M.H., Ibegwan C.,
RA Jaisl M., Kailush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lako P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Menkulov G., Milshina N.V., Nobayri C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skipski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner C., Venter E., Wang A.H., Wang X.,
RA Wang Z.Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodgerger, Morley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zhang X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195 (2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Celniker S.E., Adams M.D., Krommiller B., Wan K.H., Holt R.A.,
RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA Ferreira S., Frise E., Galie R.F., Garg N.S., George R.A.,
RA Gonzalez M., Houck J., Hoskins R.A., Hoellin D., Howland T.J.,
RA Ibegwan C., Jaisl M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA Pacle J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA Phuanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.,
RT "Sequencing of Drosophila melanogaster genome."
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Miara S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.B., Smith C.D.,
RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
RA Clump M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA Krommiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA Searle S.M.J., Smith E., Shu S., Smutnack F., Whitfield E.,

```

RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT "Annotation of Drosophila melanogaster genome."
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RA Player; ;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; A0003495; AAN09587.1; -;
SQ SEQUENCE 3432 AA; 349491 MW; FE907ACT5578657A CRC64;

Query Match 12.9%; Score 968.5; DB 5; Length 3432;
Best Local Similarity 32.8%; Pred. No. 2.1e-44;
Matches 405; Conservative 76; Mismatches 495; Indels 259; Gaps 64;

QY 111 TSPSSSKAPPSGAS---QTIKSTKRSKPPNKKTKVISEEITEHSENOESS 167
DB 2229 TAKPTTLK--PTBGSAQPTTLKPTGTTAKPTTLKPTGTT--SAKPTTLKPTGTTAKPT 2285
QY 168 SSSSSSSSSSTWIKSSKNSAANRELQKKLVKONKKNRTKKKPTPKPVVDEAGSGLD 227
DB 2286 TLKPTGTSKAPPTTLKPTGTTA---KPTTLKPTGTRTTAKPTTLK--- 2330
QY 228 NGDFVTTTBDTS---TTOHNVSTSPKITAKP-----INPRSLPBNDSKESNL--- 276
DB 2331 -----TEGTSKAPPTTLKPTGTSKAPPTTLKPTGTTAKPTTLKPTGTTAKPTTLKPT 2383
QY 277 --TVNKETVE--TKETT---TNNKOT--STDGKEKT-----TSAKET--QSIKETSAN-- 319
DB 2384 EGTSAKAPPTTLKPTGTTAKPTTLKPTGTSKAPPTTLKPTGTTAKPTTLKPTGTSKAPT 2443
QY 320 DIAPTSKVLAKPT--PKAETTTKGPALTTPEPTTPKEPASTTPKEPT---PTTKS 373
DB 2444 TLKPTGTTAKPTTLKPTGTTAKPTTLKPTGTTAKPTTLKPTGTTAKPTTLKPTGTTAKPT 2499
QY 374 APTTKEPA---PTTKSAPTTPKEAPPTTKEPA---PTTKEPAPTTKEPAPTTTS 427
DB 2500 TEGTTAKPTTLKPTGTSKAPPTTLKPTGTTAKPTTLKPTGTSKAPPTTLKPTGTTAK-- 2558
QY 428 APTTKEPAPTTPK---KPAPTTPEAPPTTPKEPT---PTTKEPAPTTKEPAPTTPK 480
DB 2559 -PTTLKPTGTTAKPTTLKPTGTSKAPPTTLKPTGTTAKPTTLKPTGTSKAPPTTLKPT 2617
QY 481 EPAPTAKKPAAPTTPEAPPTTPEAPPTTKEPS---PTTKEPAPTTTSKAPTTKEP 537
DB 2618 EGTAKAPPTTLKPTGTSKAPPTTLK--PTGTTAKPTTLKPTGTSKAPPTTLKPTGTTAKP 2676
QY 538 A---PTTTSAPTTPEPSTPTTKEPA-----PTTKEPAPTTPK---KPA-- 577
DB 2677 TLLKPTGTSKAPPTTLKPTGTTAKPTTLKPTGTTAKPTTLKPTGTTAKPTTLKPTG 2736
QY 578 ----PTTKEPAPTTPK---EPAPTTTKEPA---PTAKPEAPPTTKEAPTT--PKKL 624
DB 2737 TSAKPTTLKPTGTTAKPTTLKPTGTTAKPTTLKPTGTSKAPPTTLKPTGTTAKPTTL 2796
QY 635 TPT-----TPEKLAAPTTPEKAPPT-----PEELAPT--TPEEPTTPEEAPPT 667
DB 2797 KPTDGTAKPTTLKPTGTSKAPPTTLKPTGTTAKPTTLKPTGTSKAPPTTLKPTGTTA 2856
QY 668 TPKAAPNTPKPAAPTTPEAPPTTPEKAPPTTPEKAPPTTPEKAPTT-----PTTL 717
DB 2857 KPTTLKPTGTSKAPPTTLKPTGTTAKPTTLKPTDGTAKPTTLKPTGTSKAPPTTL 2913
QY 718 KEPAAPTTPK---KPAKELAPTTTKEPTSTSDKPA-----PTTKEGAPPTP 762
DB 2914 KPTGTTAKPTTLKPTGTTAKPTTLKPTGTSKAPPTTLKPTGTTAKPTTLKPTGTTA 2973
QY 763 KEPAAPTTPEAPPTTPKGAAPTTTLKPEAPPTPK---KPAKELAPTTTKEPTSTSDK 818
DB 2974 K---PTTLK---PTGTSKAPPTTLKPTGTTAKPTTLKPTGTSKAPPTTLKPTGTTAKP 3027

QY 819 A-----PTTKEAPPTTPEAPPT--PKKPAPTTPEPPTTSEVSTPTTK-- 864
DB 3028 TLLKPTGTSKAPPTTLKPTGTTAKPTTLKPTGTSKAPPTTLKPTGTTAKPTTLKPT 3084
QY 865 -----EPTTIHKSDESTEPBLSAETP--KALENSPKPEGV-----PTTKPAATKPEMT 912
DB 3085 TEGTSKAPPTTL--KTEGTT---TAKPTTLKPTGTTAKPTTLKPTDGTAKPTTLKPTG 3139
QY 913 TTAOKTTERDRTTPEPTT-----AARMTKEVATTEKTESKIT--ATTQVST 963
DB 3140 TSAKPTTLKPTGTTAKPTTLKPTGTSKAPPTTLKPTGTTAKPTTLKPTGTSKAPPTTL 3199
QY 964 TQODTTPKITTL-----KTTLAKVTTTK--TITTEINNKBEETKPKDRAFN 1013
DB 3200 KPTDGTAKPTTLKPTGTSKAPPTTLKPTGTTAKPTTLKPTGTSKAPPTTLKPTD--GTT 3258
QY 1014 SKATPKPKQPKPAK---KPT--STKKPKTMP-----RVKPKTTPTRKMTSTMP--ELN 1063
DB 3259 AKPTTLKPTGTSKAPPTTLKPTGTTAKPTTLKPTGTTAKPTTLKPTGTSKAPPTTLK 3317
QY 1064 PFSRIABMLOTTBPNOTPNKSLVEVNPKSBDAGAGETPHMLRPHVEMEVTPMD 1123
DB 3318 PTE--GTTAKPTTLKPTGTSKAPPTTLK---TEGTT---AKPTTLKPTGTSK 3364
QY 1124 YLPRVNGIINPMLSDFTNINCNGKPYDGLTTLR 1158
DB 3365 PTLKPTGTTAKP-----TTL--KPTDGTTPER 3391

RESULT 9
076894
ID 076894 PRELIMINARY; PRT; 1795 AA.
AC 076894;
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
DE EG-56G7.1 protein.
OS EG-56G7.1 OR CG14796.
GN Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxId=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brannon R.C., Rogers Y.-H.C., Blazey R.G., Chapple M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abriil J.F., Abghyan A.A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Baas A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos A.V., Bernan B.P., Bhandari D., Bolishakov S.,
RA Borisy D., Botchan M.R., Bouck J., Brockstein P., Brothier P.,
RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferretti W.M., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glaeser K.,
RA Glisdek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris K.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibbegan C.,
RA Jaislin M., Kalush F., Karpen G.H., Ke Z., Kentison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Liaw P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,

QY 498 PAFTTKEPAFTTKEP-----SPTTKEPAFTTTSAPTTTKEPAFTTTSAPTTT-KEP 552
 DB 654 AENSNVVP-PTTVKVPVKMKKMPWEDDDAPAKPVLSPEBCK-TVLAKKAPTTKDSLA 711
 QY 553 SPTTKEPAFTTKEP-EPAPTTKEPKAP-TPKEPAFTTTPKEPAFTTTPKEPAFTT-PKE 607
 DB 712 AADPVGSPSSKOPKLAKKAPVKPRDPSPMKAVPIKAPKT--EVPPAVVKKKEPVAKSD 769
 QY 608 PAFTTKEPAFTTTPKLTPTT-----EKLAPTTKEPAFTTPELAPTTPEEP-----T 657
 DB 770 PEPKAPK-ABPNP--VPPPTVKNPVKMKMPWEDDDAPAEVNVNPEBCKTPVLAKKT 826
 QY 658 PTTPEEPAPTTPKAAPNTPEKEPAFTTTPKEPAFTT-PEKEPAFTTTPKEPAFTTTPKTAPTT 716
 DB 827 PVKPRDPSKAPVAPKSTKTDAPVSKPEVSPKESPKAPAPNSVPV-----PTP 882
 QY 717 LKEPAFTTTPKAPKELAPTT-TPKEPTSTSDKAPTTTPKSTAPTTTPKEPAFTTTPKEPAFTT 775
 DB 883 VKNPVKM-KPWEDEDEFTBEVKRSB--PEKKTPLAK-KEPEKPD-APKVAKPRD 937
 QY 776 TTPKSTAPTTKEPAFTT-----TPKAP-----KELAPTTTKEPT-----TTSOK 817
 DB 938 PEPKAPVPE--KEPAVNAKPRDLSPKA1P1PANTQAPPTTVKNPVKMKMPWEDDE 995
 QY 818 P-----APTTKET-----APTTKEPAFTTTPKEPAFTTTPPTTSEVSTPTTKEP 866
 DB 996 PAEPVAPPEBCKTPVLAKKAPKPRDP--SPKAAVNAK-PDKIRIV-PTTVKNP 1050
 QY 867 TTIHSP-----DESTEPISA-EP--TPKALNSPKPGV-----PTTKTPA 905
 DB 1051 VKMKKPEWEDDESPSAPPEBCKTPVLAKKAPKTATKPDSEAADPVGSPSTKODK 1110
 QY 906 AT-----KPEMTTAKKOTTERDLATTEPTT--TAAPKKTETATTEKTESKITAT 957
 DB 1111 LSKAPVEKPEPTDCKDKLKSAPKKEKAPAPAKMKVWDDDEPEADPTVA 1170
 QY 958 TQVSTTTTODTTPPKITTLTKTLAKPKVTTTKTITTTTEIMNKPESTAKPKOATNSKAT 1017
 DB 1171 PSKKPTPEPADPLG-----GKTKDPK-----LKKAAKAEKTEK----- 1206
 QY 1018 TPKPKPTAPKPKPTSTKPKTPRVRKPK-----TTPTPKMTSTMP 1060
 DB 1207 PKPKVSKPEPKPTPEPKP-AAPKMKMPWEDDEPEADPTMPAPKPKPTEDP 1259

RESULT 13
 ID Q9VVG2 PRELIMINARY; PRT; 934 AA.
 AC Q9VVG2;
 DT 01-MAY-2000 (TEMBLrel. 13, Created)
 DT 01-OCT-2002 (TEMBLrel. 22, Last sequence update)
 DT 01-MAR-2003 (TEMBLrel. 23, Last annotation update)
 DB CG13731 protein.
 GN CG13731.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OC NCBI_TaxID=7227;
 OX [1]
 RN RN
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkeley;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Suton G.G., Morten J.R., Yandell M.D., Zhang O., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer J.G., Champe M., Pfeiffer B.D.,
 RA Man K.H., Doyle C., Baxter E.G., Heit G., Nelson C.R., Miklos G.L.G.,
 RA Abri J.F., Agayuni A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Bence P.V., Berman B.P., Bhandari D., Bolshakov S.,

RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
 RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahle C., Davenport L.B., Davies P.,
 RA de Fabios B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hosien D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kuip D., Lai Z.,
 RA Lasoo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merulov G., Milehina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusser D.R., Pacle J.M.,
 RA Palazzolo M., Peltman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spredling A.C., Stapleton M., Strong R., Sun E.,
 RA Switkes R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zhang X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
 RA "The genome sequence of Drosophila melanogaster.";
 RT Science 287:2185-2195 (2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Celinker S.E., Adams M.D., Krommiller B., Wan K.H., Holt R.A.,
 RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
 RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
 RA Carlson J.M., Center A., Champe M., Davenport L.B., Dietz S.M.,
 RA Dodson K., Dorese V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
 RA Ferreira S., Frise E., Galle R.F., Garg N.S., George R.A.,
 RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
 RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
 RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
 RA Pachek J., Paragas V., Park S., Patel S., Pfeiffer B.,
 RA Phouanavong S., Peltman G.S., Puri V., Richards S., Scheeler F.,
 RA Stapleton M., Strong R., Switkes R., Tector C., Tyler D.,
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.,
 RT "Sequencing of Drosophila melanogaster genome.";
 RT Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
 RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,
 RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celinker S.E.,
 RA Clump M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
 RA Krommiller B., Marshall B., Milburn G., Richter J., Russo S.,
 RA Searle S.M.J., Smith E., Shu S., Smutnick F., Whitfield E.,
 RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.,
 RT "Annotation of Drosophila melanogaster genome.";
 RT Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Adams M.D., Celinker S.E., Gibbs R.A., Rubin G.M., Venter C.J.,
 RA Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
 RN [5]
 RP SEQUENCE FROM N.A.
 RA Flybase;
 RT Submitted (SEP-2002) to the EMBL/Genbank/DBJ databases.
 DR EMBL, AF003524; AAF49349.3; -
 DR Flybase; FBgn0036717; CG13731.
 SO SEQUENCE 934 AA; 102266 MW; BCCCF1BBB3620BE CRC64;

Query Match 11.2%; Score 843.5; DB 5; Length 934;
 Best Local Similarity 32.9%; Pred. No. 3.4e-38;
 Matches 284; Conservative 48; Mismatches 371; Indels 161; Gaps 41;

```

QY 324 TSKVLAKETPKAE-----TTKGAALTTPEKPTPTPKAPASTTPEKPTP 368
DB 161 TSHFLQKRTAAEGFVPRPDPVPLVPTTQPP--TTPRRPPTRRPPTRRPPTP 218
QY 369 TTISAPPTPKAPATTTSAPTTKAPATTKEAPATTKEAPATTKEA---PTTT 425
DB 219 TYL---PPTNKPDPVTTTLRPPPPPPPPPTRRPPT---RPPTRRPATVLPPTN 268
QY 426 KSAP--TTPKEAPATTKEAPATTKEAPATTKEPT---PTT--PKEAPATTKEAPATT 479
DB 269 KPLPPTVTRLRPPPPSPPTPPPTPPPTPPPTPPATVLPPTNKPDPVTTTLRPPPP 328
QY 480 KEAPATAPKAPATT--PKEAPATTKEAPATTKEAPATTKEAPATTKEA---PTT 532
DB 329 RPPPTRRPPTKPTPPPTPPATVLPPTNKPDPPTTRRPPPTRRPPTPPPTPPATV 388
QY 533 -TTKEAPATTKSAPTTKPEPSPTTKKPA--PTT--PKEAPATTKEAPATTKEAPATT 587
DB 389 PTNKPDPVTTVTRTTTRPPTPPPTPPPTPPPTVLPPTVTRTTRPPPPPTPPKPT 448
QY 588 T--PKEAPATTTKKAPATAPKAPATTKEAPATTKEA---PTTTPPEKAPATT 645
DB 449 TYLPPTVTRTTTRAPPTTRP---PTRP--PTVPTTRRLTTPAPVLPPTNKPDP--P 502
QY 646 BELAPTTPEBPT--PTTPEBAPATT--PKAAPNTPKAPATTKEAPATTKEA--- 698
DB 503 RTVTRTT--RPTLPPTKPTPPPTTYPPTVTRTTRPPPPPTPPPTPPPTTYPPT 561
QY 699 --TTPKEAPATTPKGTAPTTLKEAPATTPKKAPAPBELAPTTKEPTSTTSKAPATT 756
DB 562 TRAPPTPPPTPPPTPPPTPPPTPPPTPPPTPPPTPPPTPPPTPPPTPPPTPP 612
QY 757 TAPTPKEAPATTKEAPATTKEA---PTT--LKEAPATTPKKAPAPBELAPTTKGTST 815
DB 613 TKTTPR---PTLPPTKPTTRP---PTTYPPEPVTRTTRPPPTPPPT--- 655
QY 816 DKAPATT--PKEAPATTKEAPATTPKKAPATTPEPTPPPTSEVSTPTTKEPTTIH 873
DB 656 -KP--PTTYPPTVTRTTATPPPTTRP---PTRPPTVPTTRRLTTPAPVLPPT-- 708
QY 874 DESPELSAEPPLKALENSPKEPGVTTKTPAATKEMTTAKOKTERDILTTPETTT 933
DB 709 PPVTVTRTTRTPR---PTLPPTKAP--TRP--PTTYPPTVTRTTRPPPPPT-- 754
QY 934 APKMTKEATATTEKTESKITATTQVSTTQDT---PKKITLKTTLTAPKVTTK 990
DB 755 -----RPTKPTPTTYPPTVTRTTATPPPTPPPTPPPTPPPTPPPTPPPT 801
QY 991 TTTTEINMKKEBETAKPKDRAINSKATTPKQ--KETKAPKXPTSKKKTMRVAKPT 1049
DB 802 --TYLPPTNKP---LPVTVTRTTTRTTRPTLPPTTRPTKPTTYPPTVTRTTRPP 855
QY 1050 PTPKMTSTMBELNFTSIAEAMLOTTTRPNOTPNSK--LVEVNPKESEDAGAGET 1108
DB 856 PT-----RPTKPTPTTYPPTVTRTTATPPPTPPPTPPPTPPPTPPPTPP 895
QY 1109 LRPHVMEVTPDMD--YLPRVNO 1131
DB 896 VRTTVYVPPPTQRTTVYVPPAPT 919

```

RESULT 14

```

ID 08H605 PRELIMINARY; PRT; 1607 AA.
AC 08H605;
DT 01-MAR-2003 (Tremblrel. 23, Created)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Elicitor-like mating protein M81.
OS Phycophthora infestans (Potato late blight fungus).
OC Eukaryota; stramenopiles; Oomycetes; Pythiales; Pythiaceae;
OC Phycophthora.
NCBI_TaxID=4787;

```

```

RN [1]
SEQUENCE FROM N.A.
RP MEDLINE=2216232; PubMed=12180924;
RA Fadietus A.U.; Cvitanich C.; Judelson H.S.;
RT "Stage-specific gene expression during sexual development in
Phycophthora infestans."
RL Mol. Microbiol. 45:1057-1066(2002).
DR EMBL; AF507055; AAN37688.1;
SQ SEQUENCE 1607 AA; 173274 MW; 9D5A588EAD3411C CRC64;

Query Match 11.0%; Score 829; DB 10; Length 1607;
Best Local Similarity 28.8%; Pred. No. 3.8e-37;
Matches 371; Conservative 113; Mismatches 445; Indels 360; Gaps 72;

QY 99 DYESFCAEV-----HNPSPSSKAPPPSGASQTITST----- 132
DB 380 DYE--YCELDENYTTIGGEWLGHSEKHPDFLWPAKPSAGVTDGISTADVALIEE 438
QY 133 -----TKSPKPNKKTKKVVIESEITEHSVSENQSSSSSSSSSTIKIXSSK 186
DB 439 SLDCVDVTSDFPDFTSTK-----PGTDYKSPSYTKTPT----- 476
QY 187 NSAAARELOKKLVKDNKKNRTKKKPTKPPVVDAGSLDNGDFKVTTPSTTQHNV 246
DB 477 -----KSPPTKSPPTKSPPT--KPTVPPETVAPETTAPATVPPETVAPET 523
QY 247 STSPKITAPINRPSLPPNSDTSKETS-----TVNKETV-----ETKETTIN 293
DB 524 STYPEEYA--PETAAPATVPPETVSPETTAPSTVPPETVAPETTAPATVPPETV 582
QY 294 KQT---STDGKETTSA---KETQSIKTSKADLAPTSKVLAKETPKAE---TKGPA 343
DB 583 PSTVPPETVAPETTAPATVPPETVSPETTAPSTVPPETVAPETTAPATVPPET 639
QY 344 LTPK--EPPTTKEPASTTPKAP--TPTTKSAPTT--PKEAPATTKSAPTTKEAP 398
DB 640 TTAATVPPRHTHPPSSNVPATVSPET--TAPSTVPPETVAPETTAPATVPPET 697
QY 399 TTKKEAPATT--PKEAPATTKEAPATTKSAPTTKEAPATT-----PKKAPATT-- 448
DB 698 ETT---ASTVPPETVAPETT--APATVPPETVSPETTAPSTVPPETVAPETTAP 751
QY 449 KE---PAPTPKKEPTPTT--PKEAPATTKEAPATTKEAPATT-----APKAPATT-- 494
DB 752 PETVSPETTAPSTVPPETVAPETTAPATVPPETVSPETTAPSTVPPETVAPET 811
QY 495 -----PKEAPATT--PKEAPATTKEPSPTPKAPATTKSAPTTKEAPATTKSAP 547
DB 812 PETVSPETTAPSTVPPETVAPETT--APATVPPETVSPETTAPSTVPPETVAP 870
QY 548 TPKEPSPTTKEAPATT-----PKEAPATT-----PKKAPATT--PKEAP-- 590
DB 871 PPEVSPETT--ASTVPPETVAPETTAPATVPPETVSPETTAPSTVPPETVAPET 927
QY 591 EPAPTTTKKAPATAPKAPATT--PKEAPATT-----PKLAPTT-----PEK 634
DB 928 ATVPETVSPETTAPSTVPPETVAPETTAPATVPPETVSPETTAPSTVPPETVAP 987
QY 635 TT-----PKEAPATT-----PEELAPTT-----PEEPPTT-----PEE 666
DB 988 ATVPETVSPETTAPSTVPPETVAPETTAPATVPPETVSPETTAPSTVPPETVAP 1047
QY 667 TT-----PKAAPNT-----PKEAPATT-----PKEAPATT-----PKE 698
DB 1048 ATVPETVSPETTAPSTVPPETVAPETTAPATVPPETVSPETTAPSTVPPETVAP 1107
QY 699 TT-----PKEAPATT-----PKGTAPTTLK-----EPAPTTPKKAPAPKELAP 739
DB 1108 ATVPETVSPETTAPSTVPPETVAPETTAPATVPPETVSPETTAPSTVPPETVAP 1167
QY 740 ---EPSTTSKAPATT-----PKGTAPTT-----PKEAPATT-----PKEAP 775
DB 1168 ATVPETVSPETTAPSTVPPETVAPETTAPATVPPETVSPETTAPSTVPPETVAP 1227

```


Mon Dec 8 09:50:45 2003

us-09-556-246-1.rpt

Page 16

Job time : 63 secs

.....

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using SW model

Run on: December 8, 2003, 09:33:27 ; Search time 27 Seconds

(without alignments)
5000.771 Million cell updates/sec

Title: US-09-556-246-1

Perfect score: 7526
Sequence: 1 MAMKTLPIYLLLLSVFVIQ.....AAATTRSGQTLSKYWYVNC 1404

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 9616882 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1090	14.5	3020	2 A43932	mucin 2 precursor,
2	950	12.6	1664	2 T18262	S-layer protein -
3	863.5	11.5	1489	2 T31108	cyst germination s
4	846	11.2	1274	2 T16251	hypothetical prote
5	812	10.8	2187	2 T30826	nascent polypeptid
6	809	10.7	1367	1 S48478	glucan 1,4-alpha-g
7	784	10.4	1188	2 S49815	extensin-like proc
8	695	9.2	1151	2 T18535	high molecular mas
9	676.5	9.0	1229	2 T25697	hypothetical prote
10	672	8.9	1344	1 A35175	mucin 1 precursor,
11	666.5	8.9	7962	2 T38346	elastic titin - hu
12	663	8.8	3570	2 T45025	mucin MUC5B, trach
13	658.5	8.7	3507	2 T34513	hypothetical prote
14	643.5	8.6	990	2 T51618	nucleolar phosphop
15	633	8.4	489	2 T11622	extensin class 1 p
16	632	8.4	761	2 C84672	hypothetical prote
17	629	8.4	6642	2 T29757	protein UNC-89 - C
18	627.5	8.3	971	2 T19431	hypothetical prote
19	622.5	8.3	839	2 T75518	hypothetical prote
20	607.5	8.1	801	2 T29018	hypothetical prote
21	607.5	8.1	924	2 S27923	gene Lpf protein -
22	605	8.0	379	2 S50125	larval glue protei
23	594	7.9	4776	2 E95206	cell wall surface
24	591.5	7.9	2232	2 T34434	hypothetical prote
25	568.5	7.6	1630	2 A53577	asectes sitoglyco
26	566	7.5	883	2 E90577	11p0protein vacu
27	563	7.5	880	2 D89756	protein T23E7.2b
28	559.5	7.4	350	2 S22456	hydroxyproline-ric
29	559	7.4	875	2 S23760	polymorphic adhes

30	556.5	7.4	856	2 T16543	hypothetical prote
31	551	7.3	620	2 S06733	hydroxyproline-ric
32	551	7.3	1087	1 QFMSH	neurofilament trip
33	545	7.2	873	2 A47283	calphostin - fruit
34	542	7.2	369	2 S20500	hydroxyproline-ric
35	540	7.2	1459	2 T32271	hypothetical prote
36	538.5	7.2	416	2 J00465	extensin precursor
37	537	7.1	1072	1 A37221	neurofilament trip
38	532	7.1	756	2 T27642	hypothetical prote
39	530.5	7.0	813	2 S70795	vsaa protein precu
40	530.5	7.0	1162	2 JH0557	exo-alpha-sialidas
41	530	7.0	865	2 A47282	calcium-binding pr
42	522	6.9	328	2 J00985	hydroxyproline-ric
43	521.5	6.9	854	2 S02003	neurofilament trip
44	518	6.9	866	2 T45462	membrane glycopro
45	518	6.9	1611	2 T38236	hypothetical prote

ALIGNMENTS

RESULT 1
A43932
mucin 2 precursor, intestinal - human (Fragments)
N:Alternate names: mucin SMUC-41
C:Species: Homo sapiens (man)
C:Date: 10-Mar-1993 #sequence revision 12-Apr-1996 #text change 05-Nov-1999
C:Accession: A49963; A45106; B45106; A43932; B33532; A61257; P00348; P00329
R:Gum Jr., J.R.; Hicks, J.W.; Toribara, N.W.; Siddiki, B.; Kilm, Y.S.
J. Biol. Chem. 269, 2440-2446, 1994
A:Title: Molecular cloning of human intestinal mucin (MUC2) cDNA. Identification of the
A:Reference number: A49963; MUID:94132002; PMID:9300571
A:Accession: A49963
A:Molecule type: mRNA
A:Residues: 1-639 <GU1>
A:Cross-references: GB:M94131; NID:9186395; PIDN:AAA59163.1; PID:9186396
R:Gum Jr., J.R.; Hicks, J.W.; Toribara, N.W.; Roche, E.M.; Lagace, R.E.; Kilm, Y.S.
J. Biol. Chem. 267, 21375-21383, 1992
A:Title: The human MUC2 intestinal mucin has cysteine-rich subdomains located both upst
A:Reference number: A45106; MUID:93016075; PMID:1400449
A:Accession: A45106
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 626-1895 <GU2>
A:Cross-references: GB:M94131; NID:9186395; PIDN:AAA59163.1; PID:9186396
A:Note: sequence extracted from NCBI backbone (NCBIP:116706)
A:Accession: B45106
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 2037-3020 <GU3>
A:Cross-references: GB:M94132; NID:9186397; PIDN:AAA59164.1; PID:9186398
A:Experimental source: Colon
A:Note: sequence extracted from NCBI backbone (NCBIP:116698)
R:Toribara, N.W.; Gum Jr., J.R.; Culhane, P.J.; Lagace, R.E.; Hicks, J.W.; Petersen, G.J
J. Clin. Invest. 88, 1005-1013, 1991
A:Title: MUC-2 human small intestinal mucin gene structure. Repeated arrays and polymor
A:Reference number: A43932; MUID:91358717; PMID:1885763
A:Accession: A43932
A:Molecule type: DNA
A:Residues: 1345-1350, 'L', 1352-1411, 'S', 1413-1448, 'P', 1450-1503, 'T', 1505-1915 <TOR>
A:Cross-references: GB:M74027; NID:9188863; PIDN:AAA59875.1; PID:9188864
A:Note: sequence inconsistent with the nucleotide translation
R:Gum, J.R.; Byrd, J.C.; Hicks, J.W.; Toribara, N.W.; Lamport, D.T.A.; Kilm, Y.S.
J. Biol. Chem. 264, 6480-6487, 1989
A:Title: Molecular cloning of human intestinal mucin cDNAs. Sequence analysis and evid
A:Reference number: A33532; MUID:89197956; PMID:2703501
A:Accession: B33532
A:Molecule type: mRNA
A:Residues: 1916-2193 <GU4>
A:Cross-references: GB:M22405; NID:9188873; PIDN:AAA6334.1; PID:9188874
A:Experimental source: intestine
R:Fany, B.H.; Gallup, M.W.; Yan, P.S.; Gum, J.R.; Kilm, Y.S.; Baebaum, C.B.

Query Match 12.6%; Score 950; DB 2; Length 1664;
 Best Local Similarity 31.0%; Pred. No. 7.4e-22;
 Matches 312; Conservative 104; Mismatches 361; Indels 230; Gaps 55;

QY 327 VLAKPP-
 DB 758 VVQPAKASDEPIPTDPSDEPTPS-----DEPTPS-----DEPTPSDEPTPSD 804
 QY 386 TKSAPPTKKEPATTKEPATTKEPAP-----PTTKKEPATTKAPPTKKEPATTK 441
 DB 805 EPTPSSEPTPEPIPTDPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPEEPIPTD 864
 QY 442 KPAPPTKKEPATTKEPATTKEPATTKEPATTKEPATTKEPATTKEPATTKEPATTKEP 500
 DB 865 SDEPTPSDEPTPS--DEPTPS--DEPTPS--DEPTPSDEPTPEEPIPTDPSDEPTPSDEPT 919
 QY 501 TPKKEPAP--PTTKKEPAPT--TPKEPAPT--APTTKEPAPT--TTKAPPTKKEP 553
 DB 920 SDEPTPSDEPTPSDEPTPSDEPTPEEPIPTDPSDEPTPSDEPTPSDEPTPSDEPTPEE 979
 QY 554 ----PTTKKEPAPT--TPKEPAPTTPKKAPATTKEPATTKEPATTKEPATTKAPATTAPK 608
 DB 980 PSDEPTPSDEPTPSSEPTPEEPIPTDPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDE 1039
 QY 609 APT--TPKEPAPTTPKKAPATTKEPATTKEPAPT--TPKEPAPTTPKEEPTPT--TP 661
 DB 1040 TSDDEPTPEE--TPKEEPIPTDPSDEPTPSDEPTPSDEPTPSD--EPTPSDEPTPEEPT 1094
 QY 662 EEPAPTTPKAAANTPKEPATTKEPAP--PTTKKEPAPT--TPKEPAPTTPKGTAPPTTK 718
 DB 1095 EEPAPTTPDPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPEEPIPTDPSDEPTPS 1154
 QY 719 EEPAPTTPKAPAPKELAPTTTKEPT--STSDKAPATTTPKGTAPTTKEPATTKEPAPT 777
 DB 1155 EPTPS--DEPTPSD--EPTPSDEPTPEEPIPTDPSDEPTPSDEPTPS--DEPTPS-- 1208
 QY 778 PKGTAPTTTKEPAPT--TPKKAPAPKELAPTTTKEPTSTSDKAPATT--TPKEPAPTTPKE 833
 DB 1209 ----DEPTPSDEPTPSSEPTPEEPI-----PTDPSDEPTPSDEPTPSD--EPTPSDE 1253
 QY 834 PAPTTPKKAPAPT--TPETPTPEEPTSEVSTPTTKKEPTTIHNSPDESTPELSAPTPKALENS 892
 DB 1254 PTPS--DEPTPSSEPTPEEPIPTDPSDEPTPSDEPT--PSDEPTP--SDEPTPSDEPTP 1305
 QY 893 PKEPGPTTKTAPAT--KEEMTTAKOKTERDLKTTPEETTTAAPKOKTETATTTKEPT 950
 DB 1306 SDEPTPSDEPTPSSEPTPEEPIPTDPSDEPTPSD--EPTPSDEPTPSDEPTPSDEPTPS 1361
 QY 951 SKITATTGVTSTTGTDTTPPKITTKTTTAPKTTTCKITTTTEIMNKEETAPKOR 1010
 DB 1362 DEPTPSDEPTPSSEPTPEEPT-----TTTTPTPESTT-----PTSG 1396
 QY 1011 ATNSKAT-----TPKOKPTKAP--KKPTSTKKKTPMRVKKPTTTPPKM----- 1055
 DB 1397 SGSGSGSGGGGGGGGTVPTSPPTPTSKPTSTPAP--TEISEPTPSDVGAIIGEENRA 1453
 QY 1056 -----TSMPLANTSRIAEAMLOTTTRPNOTPNSKLVENVKSEBDAGAGETPHM- 1107
 DB 1454 YLKGYPDGSFRPRNITRAEAVIF-----AKLL-----GADSEYGAQASAPSPYD 1498
 QY 1108 LARPH-----VFMP-----VTPTMDVLPRVNOGI----- 1133
 DB 1499 LADTHAAAMAKATISQGLPKGYPDGTTPKDONITABEPTVVLHLYTKYKGBIENSKLA 1558
 QY 1134 ---INPMUSDETNIGNG---KPVDELTL-----RNGTLVAFRGHYFMMLSPSPS 1180
 DB 1559 TIDISNPKDD---CVGHWAQEFIEKLTISIGYISGPDGT-----FKPQN 1600
 QY 1181 PARIRIEWGIPSPIDVTFRNCBEGKTPPKD--SQYAFPTNDIKD 1225
 DB 1601 YIKRSSEV---ALINRALEKPLNGAPKLPVUNESYMAF--GDIND 1642

RESULT 3
 T31108
 Cyst germination specific acidic repeat protein precursor - Phytocanthora infestans
 C/Species: Phytocanthora infestans (potato late blight agent)
 C/Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999
 C/Accession: T31108
 R/Goeinhardt, B.
 Submitted to the EMBL Data Library, April 1998
 A/Reference number: Z20986
 A/Accession: T31108
 A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: DNA
 A/Residues: 1-1489 <GOE>
 A/Cross-references: EMBL:AF061185; NID:G3851513; PID:G3851514; PIDN:AAC72308.1
 C/Genetics:
 A/Gene: car90

Query Match 11.5%; Score 863.5; DB 2; Length 1489;
 Best Local Similarity 31.7%; Pred. No. 2.3e-28;
 Matches 368; Conservative 55; Mismatches 509; Indels 229; Gaps 51;

QY 145 TKVIESEETHEHSVENQSSSSSSSSSSSTTWIKISKNSAANRELQKLYKDKN 204
 DB 282 TPTVGYSTEETEGOHVGGYEPSEDETEAPTEGITY--VREETTAAPSE--DTTVAP 334
 QY 205 KNRKKKKPPKPPVDEAGSGLDNGDFKVTPTDSTT-----QHNKSTSPKITAKPI 258
 DB 335 REVTPYAPTEKPYDVEETTYVTBESTYAPTKSETNAPTERMAHYAHLEKCDTEVTNAP 394
 QY 259 NRPRLP-----PNSDT-----SKETSLVANKETVETKEET--TNKOT 296
 DB 395 EETTVAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEET 454
 QY 297 STDGKKTSAKETSIEKTSKOLAAPSVLAKLPKAKETTTGKALTTKEPPTTPK 356
 DB 455 VAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTY 511
 QY 357 EPASTTPKEPTTTIKSAPTTKEPAPT--TKSAPTTKEPAPT--TTKEPAPTKEP 413
 DB 512 APAEETTYEPTET--TYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYEPT 570
 QY 414 PTTKEPAPTTPKSAPT-----TPKEPAPTTPKKAPATTKEPAPTTPKEPTPT 463
 DB 571 EETTVAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEET 629
 QY 464 PKEP-----AP-----TKKEPATTKEP-----APAPKAPATTKEPAP 500
 DB 630 TYAPTEETTYAPTEETTYAPTEETTYAPAEETTYEPTPEETTYAPTEETTYAPTEETTYAP 689
 QY 501 TPKKEPAPT--TTKESPTTPKEP-----APT--TKSAPT-----TTKEP----- 537
 DB 690 TETTYAPTEETTYAPAEETTYEPTPEETTYAPTEETTYAPTEETTYAPTEETTYAPTEET 749
 QY 538 --APT--TKSAPT--TPKESPTTPKEP-----APTTPKEPAPTTPKKAPATTKEPAP 586
 DB 750 TYAPTEETTYAPTEETTYAPTEETTYEPTGETTYAPTEETTYAPTEETTYAPTEETTYAP 809
 QY 587 T--TPKEPATTTPKAPAPKAPATTTPKGTAPTTPKLPTTPKELAPTTPEKAPATT 644
 DB 810 TETTYEPTPEETTYAPTEETTYEPTPEET--TYTPEETTYAPTEETTYAPTEETTYAPTE 867
 QY 645 PRELAPT--TPREPTPTPEEPT-----APT-----TPKAAAPTTPKEPATTTPK 686
 DB 868 ETTVAPTEETTYEPTPEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTY 927
 QY 687 EP-----APTTPKEPAPTTPKELAPTTPKGTAPT--TKKEPAPTTPKKAPAPK--ELAPT- 736
 DB 928 EPTTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPAEETTYEPTPEETTYAPTE 987
 QY 737 --TTKEPTSTSDKAPATTTPKGTAPTTKEPAPTTPKEP-----APTTPKGTAPT----- 784
 DB 988 ETTVAPTEETTYAPTEETTYAPTEETTYAPAEETTYEPTPEETTYAPTEETTYAPTEETTY 1047

QY 785 -----TLKEPAPTPPKP-----APKE---LAPT--TKGPTSTSDKAP 820
 Db 1048 ASTEETTYAPTEETTYAABEETPYEETTYAPTEETTYAPTEETTYAPTEETTYAPTE 1107
 QY 821 TTPKETAPTPKEAPTPPKAPPTTETPTTSEVSTPTTKPT--TIHKSDESTP 878
 Db 1108 ETTVAPEETPEPEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTY 1166
 QY 879 ELSEPTPKA-LENSPKRG-----VPTTKPAKPEETTKAKTERDRLKTP----- 928
 Db 1167 YAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTY 1226
 QY 929 ETTTAPKMTKETATTEKTESKITATTQVSTTQDT-----PKTIT-- 975
 Db 1227 BETTAP-----TEETTYEPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTY 1281
 QY 976 LKTTTLAPKVTTKTITTTTEIMNKEETAKPKDRAT--NSKATPKP-QKEPKAPKPT 1032
 Db 1282 TEETTYAPTEATTYAPTEETPYAPTEETTYEPTGETTYAPTEETTYAPTEETTYAPMEET 1341
 QY 1033 -----STKKP-KTMRYAKPKTTPPRKMTSTMPBLANTSKIAEMLQTTTRP 1079
 Db 1342 PYEPAESTSTVSTEKPCNTEETFEPTDEPTDE--PSDEPTDEPTDEPTDLPTDEPSTP 1399
 QY 1080 --NOTPNSKLVEVNPKESEDAG 1098
 Db 1400 CDNOINGIGVENKRYNNAG 1420

RESULT 4

116251
 hypochelical protein F35A5.1 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 18-Feb-2000
 C:Accession: T16251
 R:Leimbach, D.
 submitted to the EMBL Data Library, January 1996
 A:Description: The sequence of C. elegans cosmid F35A5.
 A:Reference number: Z18485
 A:Accession: T16251
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1274 <LEI>
 A:Cross-references: EMBL:U46675; NID:q1166613; PID:q1166621; PIDN:AAB52641.1; GSPDB:GNOC
 C:Experimental source: strain Bristol NZ; clone F35A5
 C:Genetic8:
 A:Gene: CESP:F35A5.1
 A:Map position: X
 A:Introns: 1272/2

Query Match 11.2%; Score 846; DB 2; Length 1274;
 Best Local Similarity 27.6%; Pred. No. 1e-27; Indels 248; Gaps 60;
 Matches 313; Conservative 119; Mismatches 455;
 QY 114 PSSKAPPSAASQTIKSTKSP-----KPP--NKKTKKVISEETIEHSVENQES 166
 Db 185 PSPKKAASKMDHPIVPTPIKNPAAKMKKPMWEDDEVTEIEKEPRATRYVPAKKKEP 244
 QY 167 SSSSSSSSSSTWIKIKSSKNSAANRELQKLVKDNK-KNRYKK-KP-----TP----- 214
 Db 245 STSVKPVSDPSPTKVK-----PVKEPEVPPPTPIKNPTKWKMKPMWEDDEFVVEVK 294
 QY 215 KPPVVDKAGSLDND-----FKVTPPDSTQHNKVSISPKITTAKP--NRPSPSPMSD 269
 Db 295 EPPVPEKKAAPVLKKDDPAKAPADSPSKAPKVEBSSVVPPTPVKNPVKKYKPEWE 354
 QY 270 TSKSTSLTNKETVETKETTNNKOTSDGKEKTSKAKETOSIEKTSADLAPTSKYLA 329
 Db 355 VDDEPAEAVKPSAPKPKTPVLAKKEPPSSSTSSDPSKKAAPAVPRDSSPKKATPL 414
 QY 330 KPTPKAE-----TTTKGPA-----LTTKPEPTPTTP-----KEPASTPK 364
 Db 415 QADPKAQEVPTPVKNPVKKYKPEWEVDDEDVVEEVKPEAPAKKTPVLAKKEPAKQTA 474

QY 365 EP-----TP-TTIKAPITPKKEPAPTTTKSAPTTKKEPAPTTTKKPA----- 405
 Db 475 KPAISKTEPEKDPVAKRDSSPKVAKADSQAAP-TVQNVKMKRPMWEDDETPA 533
 QY 406 -----PTPKKEPAPTTTKKEPAP-----TTKSAPTTKKEP-----APTTPKKEPAPTT 448
 Db 534 DVSKPDTAKTPSLAKDPAKAPKESLKPADTKAPAKRDPSPKVAFTAPEKKTPLA 593
 QY 449 KE---PAPTPKKEPAPTTTKKEPAPTTKKEPAPTTK--EPAPTAAPKAP-----TTTKE 497
 Db 594 KKEPAGPADSKTKKEBKSKRDPSPKKAAPKVPKKEVAAVAAKKEPISKPDIAKK 653
 QY 498 PAPTPKKEPAPTTTKKEP-----SPTPKKEPAPTTTKSAPTTKKEPAPTTTKSAPTTT--KEP 552
 Db 654 AEPNSPVVP-PTPVKNPVKMKKPMWEDDADAKAVSLDEPEKK--TVLAKAPKPPDSEA 711
 QY 553 SPTTKKEPAPTTPK--EPAPTPPKKAP--TTPKKEPAPTTTKKEPAPTTTKKAPTA--PKE 607
 Db 712 AADPVSGSSGKOPPLAKKAPVKKRDPSPMKAVPIKPAKPT--EVPVAVKKDEPVAKSRD 769
 QY 608 PAPTPKKEPAPTTPKLTPTTP-----EKLAPTPEKAPTTPEBLAPTTPEEP-----T 657
 Db 770 PSPKKAAP-AEPNSP--VVPPTPVKNPVKMKRPMWEDDAPAPVAVPEPEKKTPLAKKT 826
 QY 658 PTPPEEPAPTTPKAAAPVTPKEPAPTTPKKEPAPTT--KEPAPTTPKETAPTTPGTAPTT 716
 Db 827 PVKRPDPSPKKAVPAKSTKTDAPVSKKEPSPKKEBSPKKAEPNSPVVP---PTP 882
 QY 717 LKEPAPTTPKKAPKELAPTT--TTKEPTSTSDKAPPTPKGTAPTTPKKEPAPTTPKKEPAP 775
 Db 883 VKNVVKKK-KPPWEDDDEPTEVKKPSE--PEKTPVLAK--KEPEKPRD-APVAAKPRD 937
 QY 776 TTPKGTAPTTKKEPAPTT-----TPKKAP--KELAPTTKGPTS-----TTSK 817
 Db 938 PSPKKAAPVE--KEPAAVAAKPRDLSPPKAIPIPANTQEAAPT PVKNPVKMKRPMWEDDE 995
 QY 818 P-----APTTPKEP-----APTTPKEPAPTTPKKAPPTTETPTTSEVSTPTTKKEP 866
 Db 996 PAEPVSAPEPEKPKPTDPAKAPKPRD--SPKKAAPVAAK--DDPKIPEV-PTPVKNP 1050
 QY 867 TTIHKS-----DESTPELSA-EP--TPKALENSPKKEGV-----PTTKTPA 905
 Db 1051 VKMKRPMWEDDDESEVSAPEPEKTPVLAKKAPTAATPATGDEBAADPVSGTSDPK 1110
 QY 906 AT-----KPEMTTAQDKTTERDLRTTPTT--TAAPKTKETATTTTEKTESKIRATT 957
 Db 1111 LSKKAPVEKPKPTDPAKDKLKPSPAKKPEKAPPAKCKKPVWDDDPDEPEADFTVPA 1170
 QY 958 TQVSTTQDTTPPKITTLAKTTTLAPKVTTKTITTTTEIMNKEETAKPKDRATNSKAT 1017
 Db 1171 PSKKPDTEDPADPAG-----GPKTKDPK-----LKKKAPAEKPEK----- 1206
 QY 1018 TPKEPKPTKAPKKEPTSTKPKTMRVKKPK-----TTPTRKMTSTMP 1060
 Db 1207 -PKPEVSKPEPKTEPPK--AAPKMKRPMWEDDPDEPADFTWPAKKEPTEDP 1259

RESULT 5

130826
 nascent polypeptide-associated complex alpha chain, muscle splice form gp220 - mouse
 N:Alternate names: alpha-NAC protein
 C:Species: Mus musculus (house mouse)
 C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 24-Nov-1999
 C:Accession: T30826
 R:Yoclov, W.V.; St-Arnaud, R.
 A:Title: Differential splicing-in of a proline-rich exon converts alphaNAC into a muscle-
 A:Reference number: Z20889; M0UD:96312450; PMID:8658236
 A:Accession: T30826
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-2187 <YOT>

A:Cross-references: EMBL:U04863; NID:g1666688; PID:g1666689; PIDN:AA18732.1
C:Genetics:
A:Gene: Naca
A:Map position: 10
A:Insertion: 24/1, 1996/1, 2024/3, 2050/3, 2099/3, 2142/3, 2183/3
A:Note: differential splicing converts alphanac into a tissue-specific DNA-binding acti-
c:Keywords: alternative splicing; DNA binding; transcription factor

Query Match 10.8%; Score 812; DB 2; Length 2187;
Best Local Similarity 27.4%; Pred. NO. 4.3e-26;
Matches 311; Conservative 128; Mismatches 449; Indels 246; Gaps 51;

Qy	109	NPSPSPSKKAPPSGASQITKSTTKRSBPKXPNKKTKKVLSESEITEHSEVENOESS	168
Db	746	SPT--PSSSKAPVSTGA-----PPSPGADIVETESSISSKQVPAIILBPQ	792
Qy	169	SSSSSSSSSTTWMKSSKNSAANRELQKKLYKONKORTKKKKPTPKPPVDEAGSGLDN	228
Db	793	KTEPVTA SRLSNVQSPKVDPI MSD-----VTPTSPKTSATAVPK-----	833
Qy	229	GDPKVTTPDSTTQHNK-----VSTSPKITTAKPINPSPSPNSDTSKETSL--TYNKE	281
Db	834	-----DTSATLSLKSVPAVTSLSPEKAVPAENSNEATIVPTETPLTSKNAALAAATPKE	885
Qy	282	T-----TVENKETTNNKQSTODCKEKTTSAKETQSIKTSASDADLAPTSKVLAKPTPKA	335
Db	886	TLATSIPKVTSPSPQKTPKSVSLGAPAMTSSKAT--EIASKQVSPSQ--PFEVEYLL	940
Qy	336	ETTKGPA LTPKEPTPTTPKEBPASTTPKEBPTPTTKGAPATTPKE--PAPTTGSAPTTPK	394
Db	941	QHV---PPTSPKBSVSDTLGALTSPPKPPATLAEPTTPYKSPPAASKTTPATPS	997
Qy	395	-----EBAPTTKEBPATTPKEBPATTPKEBPATTT--TKSAPT-----	430
Db	998	PEGTA VPLBIPSCSKAPTAAPKESATSSSKAPATVASKELPSGVTAVLEISLP	1057
Qy	431	-----TPKEBPATTPK--PAPTTKEBPATTPKEPTPTTPKEBPATTPKEBPATTPKE	481
Db	1058	LKETSKSATPPEKSAASEPKRSPTKAGPKE--TPPGGVTA VPEISLPKPTPONATPNE	1114
Qy	482	PAPAPKAPAPATT--PKEBPAP-----TPKCEBPATTPKEBPTTPK-----EBAPTTKSA	530
Db	1115	SLAASSQKRSKTSVPKETPPGCVTAMPLEIPSAPOKAPKTA VVKQJPTPEDAVTIIAGS	1174
Qy	531	PTTKEBPATTT--KSAPTPK-----BPSPTTKEBPATTPKEBPATTPKKAAPT--	580
Db	1175	PLSPKKAASKTAAPKEAPATPSGVIAVSGEISPEPKTSKNAAPKENSATILPBRSPKTA	1234
Qy	581	-PKS-----PAPTPKEBPAP-----TTKKAPATAPKEBPATTPKEATPPT	620
Db	1235	APKETPATSEGVTA VPEISPSPTTPASKGVPTLLTPKGA PNALAE--SPASPKVPEKTA	1293
Qy	621	PKJLPTTPPEKLAPTTPEK--PAPTTPEELATTPBEPPT-----PEEPAPTPKA	671
Db	1294	APBETSTTP-----SPOKLIPK VAGPEKASATPSKTIPTKAVKETTSABEGVTAVPLE	1347
Qy	672	AAPTTPKEBPATTPKE--PAPTTPKCEBPATTPKEATPPTPKGTAPTTLKEBPATTP	725
Db	1348	IPSPRKA PKTAAPKETMPSPS--PEGATTA VQJIPSPRKSQKASGKE--TPTTPSEGV	1404
Qy	726	-----KKPAPKELAPTTKEPTS-----TSDKPA LPTPKGAPATTPKE	764
Db	1405	TAALELIPSSSKTSKMASPKETLVTPSSSKLSQTVGPKETSLGATA VPLAIPPSHKA	1464
Qy	765	PAPTPPK--PAPTPPKGTAPTTLKEBPATTPKEKAPKELAPTTTKGPTSTSDKPAPTP	823
Db	1465	PKTVDPKQVPTSPK--DAPTTLAE--SPSSPK--APKTAAPSPSR--VTVAPEKPA--TPQ	1519
Qy	824	KETAPTPKEBPATTPKKAAPTTPETP-----PPTSEVSPTPTTK-----BPTTHKS	872
Db	1520	KASGTTASKVVPVPAETGEVAASRETPVTA VPPVVKPSSHSKTSKIELKEAPATILPSP	1579
Qy	873	PDBSTPELSAEPTRALENSPKE--PGVPTTKPATATPEMTTAKDITTERDLTTEPTT	931

[illegible]

RESULT 6
 548478
 glucan 1,4-alpha-glucosidase (EC 3.2.1.3) - yeast (*Saccharomyces cerevisiae*)
 N|Alternate names: extracellular glucosylase; mucin-like protein MUC1; protein YIR019c
 C|Species: *Saccharomyces cerevisiae*
 C|Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text_change 12-Nov-1999
 C|Accession: S48478; A26877; S27281; JG6123
 R|Rowley, K.
 submitted to the EMBL Data Library, October 1994
 A|Reference number: S48478
 A|Accession: S48478
 A|Molecule type: DNA
 A|Residues: 1-1367 <ROM>
 A|Cross-references: GB:247047; EMBL:238061; NID:g603997; PID:g763364; GSPDB:GN00009; MIMF:R15447
 R|Yamashita, I.; Nakamura, M.; Fukui, S.
 J. Bacteriol. 169: 2142-2149, 1987
 A|Title: Gene fusion is a possible mechanism underlying the evolution of STA1.
 A|Reference number: A91631; MUID:87194600; PMID:3106330
 A|Accession: A26877
 A|Molecule type: DNA
 A|Residues: 1-242 <YAM>
 A|Cross-references: EMBL:M16164; NID:g172522; PIDN:AAA35014.1; PID:g172525
 A|Accession: B26877
 A|Molecule type: DNA
 A|Residues: 762-1331 <YA2>
 A|Cross-references: EMBL:M6165; NID:g172523; PIDN:AAA35015.1; PID:g172526
 R|Pardo, J.M.; Iñez, E.; Zalacain, M.; Claros, M.G.; Jimenez, A.
 FEBS Lett. 239, 179-184, 1988
 A|Title: Similar short elements in the 5' regions of the STA2 and SGA genes from *Saccharomyces cerevisiae*
 A|Reference number: S27281; MUID:89031230; PMID:3141213
 A|Accession: S27281
 A|Molecule type: DNA
 A|Residues: 1-31 <PAR>
 A|Cross-references: EMBL:X13857; NID:g4551; PIDN:CAA32069.1; PID:g4552
 R|Lambrecht, M.G.; Bauer, F.P.; Marmur, J.; Pretorius, I.S.
 Proc. Natl. Acad. Sci. U.S.A. 93, 8419-8424, 1996
 A|Title: Muc1, a mucin-like protein that is regulated by Msa10, is critical for pseudohyphal growth of *Saccharomyces cerevisiae*
 A|Reference number: JG6123; MUID:96323237; PMID:8710886
 A|Accession: JG6123
 A|Status: nucleic acid sequence not shown
 A|Molecule type: DNA
 A|Residues: 1-1367 <LAM>
 A|Cross-references: GB:U30626; NID:g1304386; PIDN:AAC49609.1; PID:g1304387
 C|Genetics:
 A|Gene: SCD:MUC1; STA2; MAL5; DEK2; SGD:S0001458
 A|Cross-references: MIPS:YIR019c; SGD:S0001458
 A|Map position: 9R
 C|Superfamily: yeast glucan 1,4-alpha-glucosidase homolog; glucan 1,4-alpha-glucosidase
 C|Keywords: glycosidase; hydrolase; polysaccharide degradation; transmembrane protein
 F|5-21|Domain: transmembrane #status predicted <TM1>
 F|1350-1366|Domain: transmembrane #status predicted <TM2>
 Query Match 10.7%; Score 809; DB 1; Length 1367;
 Best Local Similarity 27.5%; Pred. No.3.6e-26;
 Matches 328; Conservative 112; Mismatches 526; Gaps 53;
 85 DCDAGCKRDK---CCPDYESCFAEYHNPTSPSSKAPPPSGASQSTIKTSRKPFPN 141

```

Db      174 DLSTGNNYDNGHSGQDTPGPGYWNIDCDNNGCGKSSSTTSSSES--STT----- 223
Qy      142 KKKTKKVESEITEHEHVSVENQSSSSSSSSSTTWKIKSKNSKANDELQKLVK 201
Db      224 ---TSSTSSSTTSSSTSSSTTSSSTSSSTSSSTT-----APATPTTSSCTKEKPTP 274
Qy      202 DNKKRRTKKKPTPKPVVDEAGSLDNGDFKVTTPDSTTQHNKYSTPKITTAKINPR 261
Db      275 PTTTSSCTKEKPT--PPHND-----TTCCT-----KKKTTTSTCTKKTTPV 314
Qy      262 PSLPNSDTSKETSJLVNKETVEKETTNNKQSTDGKETSKEKOSIEKTSADL 321
Db      315 PT--SSSSTSSS-----APVTPSSSTT-----SSSAPVTSSTSSS--- 353
Qy      322 APTSKVLAKPTPKAETTTKGPALTTKPEPTTTPKEPASTTPEKPTTITKSAPTPKPEP 381
Db      354 APV-----PTSSSTSSSAPVTSSTSSSAPVTSSTSSSAPVTPSSSTSSS 407
Qy      382 APTT-----TKSAP-----TTPKEPAPTTTKEPAPTTTKEPAP-----TTPKEPAPTTT 428
Db      408 APVTSSTSSSAPVTSSTSSSAPVTS-----TTESSAPVTSSTSSSAPVTPPTS 463
Qy      429 PTPKEPAPTTPKKAPTTPKKAPTTPKKAPTTT-----TTPKEPAPTTPKKAPTTPKKAP 485
Db      464 STTESSSAPVT-----SSTTESSAPV-----PTSSSTTESSSAPVT-----SSTTESSAPV 512
Qy      486 APKAPAPTTPKKAPAPTTPKKAPAPTTKKSPTPKKAPAPTTKKAAPTTPKKAAPTTPKKA 545
Db      513 -----PTSSSTSSSAPAPVTS-----SSTTESSSAPVTS-----TTESSSAPVTPPTS 559
Qy      546 PTPKEPAPTTTKEPAPTTTKEPAPTT--TPKKAPAPTTTKEPAPTTTKEPAPTTTKEPAP 604
Db      560 STTESSSTPTVTS-----TTESSSAPVTPSSSTTSSSAPVPT-----PSSSTTESSSAPA 611
Qy      605 PKEPAPTTPKETAPTTPKKLTPTTPEKLAPTTPEKAPTTPELAPT--TPEEPTPTTEE 663
Db      612 PTPSSSTTESSSAPVT-----SSTTESSAP--VPTSSSTTESSSAPVTPSSSTTESSSA 666
Qy      664 PAP-----TTPKAAAPNTPKKAPAPTTPKKAPAPTTTKEPAPTTTKEPAPTTTKEPAP 717
Db      667 PVPPTSSSTTESSSAPVT-----SSTTESSSAPVT-----SSTTESSSAPVTPPTSSTESS 718
Qy      718 KEAPAPTTPKKAPKELAPTTTKEPSTSDKAPAPTTTGPATPTTKEPAP-----TTPKEPA 774
Db      719 SAPVTPSSSTTESSSAPVTPPS--SSTTESSAPVTSSTTESSSAPVTPPTSSTTESSSA 777
Qy      775 PT--TPKGAAPTTLKEPAPTTPKKAPKELAPTTTGPATPTTSDKAPAPTTTKEPAPTTTKE 833
Db      778 PVPPTSSSTTESSSAPVTPSSSTTESSVAPVPTSSSSNITSSAPSSPTSSSTESSSV 837
Qy      834 PAP-----TTPKKAAPTTPETPTTSEVSTPT-----TTPKEPTIHKSPDESPTPELSAE 883
Db      838 PVPPTSSSTTESSSAPVTSSTTESSVAPVPTSSSSNITSSAPSI---PSSSTTESPT 894
Qy      884 PTPKALESPKEGVPPTTKTPATKPEMTTAKDKTERDLRTPEETTAAPKOKTERAT 943
Db      885 GT--TTPSSSKTPGSGTETSVSSTETTTVPKTTTSVTTPSTTTTITVCSGTNSAGE 953
Qy      944 TTEKTESKITAT--TTQVSTTTQDTPPFKITLTKTTTLAPKVT---TKKITTTT--EI 997
Db      954 TTSGSPKTVTTTPTTTTTSVTSTTTTITVCSGTNSAGETTTSGCSPKITTTTTPC 1013
Qy      998 MNKPEETAKPKORATNSKATTPKPKP-----TKAPKPTSTYKPKPTM 1040
Db      1014 STSBEITA-----SSTTTSPTTPTVTVSTVTVTEYSTYKPGSEITTTTPTVKNI 1065
Qy      1041 PRVKKKPTTPPKKMTSMPBELNPTSRJAEMLQTTTPNQTNPNSKLVAVPKSDAGDA 1100
Db      1066 PTTTTLTTAAPT--SVTTVNTFTPTT--TTTVCSGT-----NSAGETTSGC 1109
Qy      1101 EGEPHMLLRHVMEVETPMDVLPVAVNOGIIINPMLSDETNINCNGKPVGLTTLRNG 1160

```

```

Db      1110 SPKTVTTTVCSTGTEGTEATLTVTAATTVTVTESSTGTNSA--GKTTGYTKSV 1168
Qy      1161 TLVAFRGHYFMNLSPSPSPARITBVGJPSIDVFTRC-----NCEGKT 1208
Db      1169 TT-----YVTTLAPASAPVTPATN-----AVPTTTT--TECSATNAAGET 1207.

RESULT 7
S49915
extensin-like protein - maize
C:Species: Zea mays (maize)
C:Date: 05-Mar-1995 #sequence_revision 12-May-1995 #text_change 29-Oct-1999
C:Accession: S49915
R:Rubinstein, A.L.; Broadwater, A.H.; Lowrey, K.; Bedinger, P.A.
submitted to the EMBL Data Library, June 1994
A:Description: Rex genes: pollen-specific genes with extensin-like domains.
A:Reference number: S49915
A:Accession: S49915
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1188 <RUB>
A:Cross-references: EMBL:Z34465; NID:9600117; PIDN:CAA84230.1; PID:9600118

Query Match          10.4%; Score 784; DB 2; Length 1188;
Best Local Similarity 29.9%; Pred. No. 3.4e-25;
Matches 230; Conservative 59; Mismatches 359; Indels 122; Gaps 33;

Qy      212 PTPKPVVDEAGSLDNGDFKVTTPDSTTQHNKYSTPKITTKAPINPRSL---PPN 267
Db      460 PTPHSPPAD-----DVPPTPPVPGKSPATSPSPQVPPASTPPSLVLKLSPPQ 510
Qy      268 SDTSKERSLVNKEETVEKETTNNKQSTDGKETSKEKOSIEKTSKDLAPTSKV 327
Db      511 APVG-----SPPPVKTTSPAPIG-----SPSPPPVSV 541
Qy      328 -----LAKPTPKAETTTKGPALTTKPEPTTTPKEPASTTPEKPTTITKSAPTTPKEPA 382
Db      542 SPPEPVSPPPAPVAGSPPEKSPPPAPVASSPPPVKSP--PPPLVASSPPPVKSP 599
Qy      383 PTTKSAAPTTP--KEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 439
Db      600 PPAPVASSPPPVKSPPTPVASPPPPAPVASSPPPKSPPPPVSSPPPEKSPPPP 659
Qy      440 PKKAPAPTTKEPAPTT-----PKE-----PPTTTPKEPAPTTTKEPAPTTTKEPAPTTAPK 490
Db      660 PAKSTPPEEPTPTPTSVKSSPPPEKSLPPTLLPSPPQEKPTPEPTPSKP--PSSEKP 718
Qy      491 APTTKEPAPTTPKKAPAPTTTKEBSPPTTTPKEPAPTTTKSAPTTPKEPAPTTTKSAPTTPK 550
Db      719 SP--PKBVSPPQTP-----KSPPPAPVSSPPTPVSSPPALAPVSSPEVKSP----- 768
Qy      551 EBSPTTKEPAPTTPKKAPAPTTPKKAPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAP 610
Db      769 PPAPVSSPPAPQVKSPPPVQVSSP--PPAPKSSPPAPVSSPPQVKEKTSPPPPAPLSPP 827
Qy      611 TTPKEAP-----TTP-----KKLTPTPEKLAAPTTPKEAP-----TTPPELAPTTPEEPT 657
Db      828 LAPKSSPPHVVSSPPPVKSSPPPPAVSSPPLTPKASPPAHVSSPPEVKESTP--PA 885
Qy      658 PTPTEBAPPTPKKAANTPKK--PAPTTPKKAPPTTTPKEPAPTTTKEPAPTTTKEPAPTTT 716
Db      886 PTTTISP--PSSEKSPPTPTVSLPPIVKSPPAPAMVSSP--PMTPKSSPPVVVSSPPT 943
Qy      717 LKE--PA-----PTPKK--PAKELAPTTTKE-----PTSTSDKAPAPTTTKEPAPTTT 762
Db      944 VKSSPPAPVSSPPAPVSSPPAPVNLPPPEVSSPPPTPVSSPPPA--PKSSPPAP 1000
Qy      763 --KEPAPTTPKKAPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 821
Db      1001 MSSPPPEVKSPPPPAVSSPPPVKSPPPAPVSSPP--PPVKSPPPPAPVSSPPPV 1057
Qy      822 --TPKEAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTT 880

```

Db 1058 KSPPPAPISPPPPVSKSPPPAPVS--SPPPVKSPPPPAPVSPPPPPKSPPPAPVS 1115
Qy 881 SAEPTKALENSPKERGVPPTTKPATKEMTTAAKTERDARTTPT 930
Db 1116 SPPAP-----VKPSLP--PPAPVSPPPVTPAPPKGEOSLPPPAS 1158

RESULT 8
T18535
high molecular mass nuclear antigen - chicken (fragment)
CISpecies: Gallus gallus (chicken)
CDate: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
CAccession: T18535
J.Shimada, K.; Harata, M.; Mizuno, S.
J. Cell Sci. 110, 3031-3041, 1997
A>Title: A nuclear matrix-associated high molecular mass nuclear antigen, hNNA, of chick
A.Reference number: Z18955; MID:9803440; PMID:9365273
A.Accession: T18535
A.Status: preliminary; translated from GB/EMBL/DBJ
A.Molecule type: mRNA
A.Residues: 1-1151 <SH1>
A.Cross-references: EMBL:D88440; MID:d1177138; PID:d1025045; PIDN:BA24137.1

Query Match 9.2%; Score 695; DB 2; Length 1151;
Best Local Similarity 25.3%; Pred. No. 1.4e-21;
Matches 304; Conservative 108; Mismatches 466; Indels 324; Gaps 59;

Qy 105 AEVNPSPSSSK-----APPASQTIKSTKSPKPKPKKKTKKVISSEETEEH 158
Db 39 ARIH--PAAPQPPKVPVIGAPPPPTGTEPT-----PSKPDGADAAPKASAEITSP 90
Qy 159 SVSENOSSSSSSSSSTIWKIKSKSAANRELQKKLVKDNKGNKTKKKPPKPV 218
Db 91 PASPPDPKAPSGAGAEAGTPPSQAG-----TPP-- 126
Qy 219 VDEAGSGLDNGDPFKVTPPTSTTQHNKVS--PKITAKPINRPSLPNSDTSKETS 276
Db 127 -SGAAGAPKGDGTAGPSGTSGADGKPAAGVPAKTTA--ATEAR-----ASAAP 176
Qy 277 TYNKETTVEKETTNN--QTSQKKEKTSNAKEQSIKTSKADLAFTSKYLAKPTK 334
Db 177 TVPKATAEATAVTAASQSAKATDMAVTA--SQSAAPKATV--EVKPAAMAANAEEAKA 232
Qy 335 AATTKGALTPPKETP--TPKESASTPKP-----TPTTKSATTKKEPAPT----- 385
Db 233 VTAAAPPAATBAKAPVTSPTIPCSAABAKELTAASPTASA--TAAKAPVATASIM 290
Qy 386 -----TKSAPTTKKEPAPTTK-----BPAPTTKKEPAPT 415
Db 291 ATKVTAABAPSPSVKATTTDKATATAPKAGPVKAVAVCAEAKAPPPPOQLK 350
Qy 416 TKKEPAPTTKSAPT-----PKESA-----PTPKKAPPTKKEPAPTTKEPT 462
Db 351 AAAAAATCTELKPAAPPHSGPSRANSHVTVTPPVPAATAV--TAGAVPKASTGT 408
Qy 463 T-----PKERAPTTKKEPAPTTKKEPAPTAAPKAPTTKKEPAPTTKEP--APTTKES 515
Db 409 TPAAPPOQVP--KAAPVTPSPPOQAVPRAATAA--APTPPOQVTKAATTNATP 461
Qy 516 PTPKKEPAPTTKSAPTTKKEPAP--TTKSAPTTKKESPTTKEP-----APTTPKEP 568
Db 462 PQQIPKATTTTATVTPOQRIPKAGTDAAPPAPVKAAPSGRATPQVPAANDPQK 521
Qy 568 APTTPKKEPAPTTKKEPAP-----TPKKEPAPT--TKKAPAPTAAPK-----EPAPT 611
Db 522 PTPPOVSPAVTEPKPAPAPPPSPNEATPAVSPSPNLSPLPTPKFVPLMALTPQV 581
Qy 612 TRK--ETAPTPPKULTPTPEKLAPTTEKAPATTPELA-----PTP----- 653
Db 582 TGAQMTQLAATKSPVVPKASPK--ALMTPPPPPGI--PRLAALAKLGLPSSVVASAMAK 640
Qy 654 --EAPPTTP--EAPAPTTP-----KAAAPNTPE-----PAPTTPKEP 688

Db 641 VTPRLPASPVMASPASLGPDAARVALATNAASPGAPBAAGNGTLMAPGAANTQM 700
Qy 689 APTTKKEPAPTTKKEPAPT-----TPKGTAPT----- 715
Db 701 AFIGAAGAAQTAPKMAATHVSPMGAGATQMSPTGAANTHNSPIGAGATQMSPMGAAN 760
Qy 716 TLKEPAPTTKKAPEKELAPTTKEP-----TSTSDKRAPPTPKG--TAPTTKEP 765
Db 761 TOMSPMGATTTQMSPMGAATTPSPMGAAATQVNTASGNMTQVSPMGAAATPQTPSG 820
Qy 766 APTTPKKEPAPTTKKEPAPT-----APTTPKKEPAPTTKKEPAPTTKKEPAPTTKKEP 818
Db 821 AATTP--GSPM-----GAATTLMSPMGAATTPG--PSPMGAVTTPPMAATNTQPPPM 872
Qy 819 APTTKET--APTTPKEP--APTTPKAP--APTTPPP-----PTSEVSTPTTK 864
Db 873 AASTQSTPMGAATTTQSPMGATTTQSPMGASTPOAPPTVAGSPTPPPPIPSPTAQT 932
Qy 865 EPTTIHKSPEESTPEL--SABPTKALENSPKERGVPPTTKPATKEMTTAKOKTER 922
Db 933 SPQMSKSPPPPPAPSAQAOTSPAANVANAPOV--TAVSPA--PIGVTAASPSADGA 988
Qy 923 DLRTTPETTAPKMT--KEATTTKTESKITATTQVSTTQDTTPPKITTKTTL 981
Db 989 RLSPPGTAATDGPASPAATDVEATD--VTAATAVPA-----EAP-----TK 1033
Qy 982 AKVTTTKKTIITTTIANKPEETAKPKDRAINSKATTPPKQKPTAPKPKTSTKKPTMP 1041
Db 1034 AKRSSSSSSSSSSSSSSSSSSSSSSSDSSSSSESNPASP-----P 1078
Qy 1042 RVRKPKTTPPKMTSTPELNPTRIAEAMQTTTRPQNTSNLVEVNPKEPDAGAE 1101
Db 1079 AVGDQOQMTPGAASVPP-----VTEAAVQ-----EAAAAAAGAE 1117
Qy 1102 GE 1103
Db 1118 RE 1119

RESULT 9
T25697
hypothetical protein Fl6F9.2 - Caenorhabditis elegans
CISpecies: Caenorhabditis elegans
CDate: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
CAccession: T25697
R.Fulton, B.
submitted to the EMBL Data Library, August 1996
A>Description: The sequence of C. elegans cosmid Fl6F9.
A.Reference number: Z20071
A.Accession: T25697
A.Status: preliminary; translated from GB/EMBL/DBJ
A.Molecule type: DNA
A.Residues: 1-1229 <FUT>
A.Cross-references: EMBL:U67956; PIDN:AA807691.1; GSPDB:GN00028; CESP:Fl6F9.2
A.Experimental source: strain Bristol N2; clone Fl6F9
A.Genetics:
A.Gene: CESP:Fl6F9.2
A.Map position: X
A.Introns: 35/1; 361/1; 384/1; 482/2; 517/1; 971/1; 1021/2; 1179/3

Query Match 9.0%; Score 676.5; DB 2; Length 1229;
Best Local Similarity 27.2%; Pred. No. 8.8e-21;
Matches 330; Conservative 95; Mismatches 385; Indels 405; Gaps 56;

Qy 73 RCFSEFERGBCDDAOCKKYDKCCPDYESPCAENVNFTSPSSSKAPPSGASQTIKST 132
Db 34 RDKFIVKHIKKQNTCTSC--KCVPD-----APSN-----PFDVSTTSS-- 70
Qy 133 TKRSPKPNKKTKKVISSEETEEHSVSENOSSSSSSSSSSSTIWKIKSKSAANR 192
Db 71 -----INNDVNDIGSPGSDSNPTGSMQIEIATVGGQTVKSEHNIDSSV 114
Qy 193 ELQKKLVKDNKGNKRTKKKPTPKPVVDEAGSGLDNGDPFKVTPPTSTTQ--HNKVS 249

A>Title: A novel core protein as well as polymorphic epithelial mucin carry peanut agglu
A.Reference number: JX0235; MUID:93123189; PMID:1478919
A.Accession: XM0066
A.Molecule type: mRNA
A.Residues: 998-1011, 'ES', 1014-1017, 1018-1032, 'T', 1034-1037, 1038-1057 <MAS>
A.Experimental source: gastric carcinoma cell
R.Zirhan-Dicht, S.; Baruch, A.; Broy-Stein, O.; Kedar, I.; Wreschner, D.H.
FEBS Lett. 356, 130-136, 1994
A>Title: Tyrosine phosphorylation of the MUC1 breast cancer membrane proteins cytokine
A.Reference number: S51026; MUID:95080414; PMID:7988707
A.Accession:
A.Note: undetermined tyrosine residues in the carboxyl-terminal non-repetitive region
A.Comment: This protein is length polymorphic. Individuals may have between 21 and 125 c
partial repeats. The repeat shown is defined by SmaI nuclease sites.
C.Comment: Serine and threonine residues in the tandem repeat domain are extensively gly
C.Comment: For an alternative splice form without a tandem repeat domain, see PIR:S48146
C.Genetics:
A.Gene: GDB:MUC1; PUM
A.Cross-references: GDB:120705; OMIM:158340
A.Map position: 1q21-q23
A.Introns: 20/1: 62/3; 1165/3; 1184/2; 1230/1; 1270/3; 1320/3
C.Superfamily: polymorphic epithelial mucin
C.Keywords: alternative splicing; duplication; glycoprotein; phosphoprotein; polymorphic
P.1-1344/Product: mucin 1 precursor, splice form A #status predicted <PRA>
P.1-62/Region: mucin 1 amino-terminal non-repetitive
P.1-23/Domin: signal sequence #link PRA #status predicted <SIGA>
P.1-19, 29-32/Domin: signal sequence #link PRA #status predicted <SIGB>
P.1-19, 29-32/Product: mucin 1 precursor, splice form B #status predicted <PRB>
P.1-19, 29-32, 1033-1344/Product: mucin 1 precursor, epithelial tumor antigen splice form
P.138-1017/Region: 20-residue repeats (GSTAPAHGVSADTRAP)
P.1143-1344/Region: mucin 1 carboxyl-terminal non-repetitive
P.1245-1272/Domin: transmembrane #status predicted <TRM>
P.1046, 1064, 1118, 1144, 1222/Binding site: carbohydrate (asn) (covalent) #status predicted
P.1213/Binding site: phosphate (tyr) (covalent) #status predicted
Query Match 8.9%; Score 672; DB 1; Length 1344;
Best Local Similarity 27.8%; Pred. No. 1.5e-20;
Matches 310; Conservative 78; Mismatches 472; Indels 256; Gaps 56;
QY 276 LVNKKETVETKETTNNK-QTSDGKEKTSKTSQSIKTSKAD-LAPTSKVLAKPT 332
DB 15 LVNLTATTPAKPAVTVTSGHASTPGCEKETSATORSVPSSTENAVSMTSVLSHS 74
QY 333 P-KAETTTGP--ALTTPKEP-----TPTTPKEPASTTP-----KE 365
DB 75 PGGSGSTTGQDVTTLAPATBPASGSAATMGQDVTSPVTRPALGTTTPAHGVTSAIDNK 134
QY 366 PTPPT-----IKSAPTTKEPAPPTT-----KSAPTTKEPAPTTKEP-----A 405
DB 135 PARGSTAPPAHGVTSAPDT--RAPGSTAPPAHGVTSAPDT--RAPGSTAPPAHGVTS 190
QY 406 PTPPKPAPPTTKEPAPTTKSAPTTPKEPAPPTP-----KKAPPTPKP----- 451
DB 191 PPT--RAPAGSTAPPAHGVTSAPDTTPRAPAGSTAPPAHGVTSAPDTTPRAPAGV 247
QY 452 --APTPKEPTPTTP-----KEPAPTTKEP-----APTPKEPAPTPAK----- 488
DB 248 TAPADTPRAPAGSTAPPAHGVTSAPDTTPRAPAGSTAPPAHGVTSAPDTTPRAPAGV 307
QY 489 KPAPTTPKEPAPPTPK-----EPAPTTKEPSPPTTPKEPAPPTT-----TKSAPTTKEPAPPT 541
DB 308 TAPADTPRAPAGSTAPPAHGVTSAPDTTPRAPAGSTAPPAHGVTSAPDTTPRAPAGV 367
QY 542 TKSAPTTKEPSPPTTTPKEP-----APTPKEPAPPTPK-----KPAPTTPKEPAPPTPK----- 590
DB 368 T-SAPDT--RAPAGSTAPPAHGVTSAPDTTPRAPAGSTAPPAHGVTSAPDTTPRAPAGV 424
QY 591 ---EPAPTTTKKAPAPAPKEPAPPTTKEPAPTTPKKLTPTTPKALPTTPKEPAPPTTP----- 646
DB 425 HGVTSAADTPRAPAGSTAPPAHGVTSAPDTTPRAPAGSTAPPAHGVTSAPDTTPRAPAGV 484
QY 647 ---ELAPTTPEBPTPTTP-----EBAPV--TTPKA-----AAPNTPKPAPPTPK----- 686

DB 485 HGVTSAADTPRAPAGSTAPPAHGVTSAPDTTPRAPAGSTAPPAHGVTSAPDTTPRAPAGV 544
QY 687 ---EPAPTTTPKEPAPPTTKEPAPTTTPKGTAPPTTLKAPAPPTTP-----KKAPAPKEL 733
DB 545 HGVTSAADTPRAPAGSTAP-----PAHGVTSAPDTTPRAPAGSTAPPAHGVTSAPDTTPRAPAGV 600
QY 724 APPTTKEPTSTSDPAP--TTPKGTAPPTTPKEPAPPTTPKEPAPPTTPG-----TAPTLK 787
DB 601 AP-PAHGVTSAPDTTPRAPAGSTAPPAHGVTS-----APDTTPRAPAGSTAPPAHGVTSAPDTTP 655
QY 788 EPAPTTTP-----KKAPAPKELAPTTTGPSTSTSDKAPAP--TTPK-----ETAPTTTP 831
DB 656 APGSTAPPAHGVTSAPDTTPRAPAGSTAP--PAHGVTSAPDTTPRAPAGSTAPPAHGVTSAPDTTP 714
QY 832 KEPAPTTTP-----KKAPAPTTPEPTTPPTTSVSTPTTKEPTT-----IHKSPD 874
DB 715 PARGSTAPPAHGVTSAPDTTPRAPAGS--TAPPAHGVTSAPDTTPRAPAGSTAPPAHGVTSAPD 772
QY 875 ESTPPLSAPPTPKALENPKPREGVPTTKTPAA-----TKPEWTTAKOK--TTERD 923
DB 773 TRPAGSTAPPAHGVTSAPDTTPRAPAGSTAPPAHGVTSAPDTTPRAPAGSTAPPAHGVTSAPD 832
QY 924 LRTTPEPT-----TAAPKMTKETATTTKTESKITAATTOVSTTODTTP--PKIT 974
DB 833 TRPAGSTAPPAHGVTSAP-----DTPRAPAGSTAPPAHGVTSAPDTTPRAPAGSTAPPAHGV 888
QY 975 TLKTTTLAPKTTTAKTTT--ELMKPEELAKKDATSKATTPKOKTPAKKPT 1032
DB 889 SAPDTTPRAPAGSTAPPAHGVTSAPDTTPRAPAGSTAPPAHGVTSAPDTTPRAP--GSTAPPAH 946
QY 1033 STKKKTPMPVRKPKPTTPTPRMTSTWELNP-----TSRIEAMLQT--TTP--NOTPN 1084
DB 947 VTSAPDTTP--APGSTAPPAHGVTSAPDTTPRAPAGSTAPPAHGVTSAPDTTPRAPAGSTAP 1003
QY 1085 SKLVEVNPKSDAGAGETPHMLLRPHVFWPEVTPDMDYLPRVNOGIIINPMUSDEN 1144
DB 1004 AHGVTSAPDTTPRAPAGSTAPPAH-----GVTSAPDNRPALGSTA-----PVHANTS 1049
QY 1145 IONGRPVUGLTLRNGTLVAFRGHFMFLSPSPS 1180
DB 1050 ASGSASGSASTLVHNGTSARATTPPASKSTPFSIPS 1085
RESULT 11
138346
elastic titin - human (fragment)
C.Species: Homo sapiens (man)
C.Date: 29-May-1998 #sequence_revision 29-May-1998 #text_change 21-Jul-2000
C.Accession: 138346
R.Label: S.; Kolmerer, B
Science 270, 293-296, 1995
A>Title: Titin: giant proteins in charge of muscle ultrastructure and elasticity.
A.Reference number: A57430; MUID:96026330; PMID:7569978
A.Accession: 138346
A.Status: preliminary; translated from GB/EMBL/DBJ
A.Molecule type: mRNA
A.Residues: 1-7962 <RES>
A.Cross-references: EMBL:X90569; NID:g1017426; PIDN:CAA62189.1; PID:g1017427
C.Genetics:
A.Gene: GDB:TTN
A.Cross-references: GDB:127867; OMIM:188840
A.Map position: 2q31-q31
Query Match 8.9%; Score 666.5; DB 2; Length 7962;
Best Local Similarity 24.2%; Pred. No. 1.4e-19;
Matches 268; Conservative 120; Mismatches 491; Indels 229; Gaps 50;
QY 98 PDYSEFCALVNPSPSPSKKAPPPSGASQITIKSTKS-----PKPPKK 143
DB 6827 PEESEFVPE--EVLPEVKPVPAVPELIKKTATEKVVIPKKEAPPAKVPEVPKV 6884
QY 144 KTKKVI--ESEEITEHVSSENOESSSSSSSSSSSTIKWIKSKNSKANRELQKLVK 201


```

Db 6885 EKKRIILKEBEVLVEVEEPEEPISEB-----EIPPEEPISEVEEVAAPRV 6935
Qy 202 DNKKRTKKKPPKPPVDEAGSGLDN--GDPKVTTP-DTSTQNKSTSPKITTAP- 257
Db 6936 EVIKKAVEAPTPVKKVEAPPAKVSKKIPEEKVAVPQKKKAPPAKVEVPKVEKVV 6995
Qy 258 -INPSPILP---NSDTSKETSLTVNKEVETK---EYTTNNKQTSIDGKE----- 302
Db 6996 LVPKKEANPAAKGRVLEBKVSVAHQEVVAKERLELVEAVEAVEIPEEPEHVEVEYF 7055
Qy 303 -----KTSAKETQSIKT-----SAKDLPSTSVLAKPTPKAETTKG 341
Db 7056 EEGEHEVEEFLKQHRVEEHRKVEKRVLEVEAEVEVEFEKPAAP-----KG 7107
Qy 342 PAIT-----TPKEPTTP-KEPASTPKPEPTTIKAPTTPKP-APTTKSAPTTPK 394
Db 7108 PEISIKIIPKKPPPTKVPKPEKPAKVEVPKVIIVEKVRVPEEPRVPTKVPEVLPPK 7167
Qy 395 EPAPPTTKEPAPPTTKEPAPPTTKEPAPTTKSAPTTKEPAPTTPKK-----PAPTPE 450
Db 7168 EVVP-----EKKVVPKAKPEAP-----PKVPEAPKEVVEPKVPPEPKKEV 7213
Qy 451 PAPTKEAPTPTTKEPAPTTKEPAPTT-----KEPAPAKKAPPTTKEPAPTT 503
Db 7214 PPTKVPKPAKAVPEKVPKPEAIPEKPEBPPEVEEPEEPEEPPAPK-KPEVPVAVPEVP 7272
Qy 504 KEPAPTTKEBPPTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 563
Db 7273 KEVPEKVPAPAP-PKKPEVTPVK-VPEAPEVVEPKVPVP-PKKPEVPTK----- 7323
Qy 564 TPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 623
Db 7324 VPEVAKVAVPEKVP-----EAIPEKPEBPPEVEEPEEVEALVEEPAVVEEPEAPQ 7379
Qy 624 LT---PTTPEKAPTT---PEKAPTTPEELAPTTPEEPTP-TTPEEAPTTKAAAP 674
Db 7380 VTPVPEKVPKPAVAVAKKELPVPVPEVPEKVPKVPVAVPKK--PEAPKAP 7437
Qy 675 NTPKEPAP-----TPKEP-----APTTPKEPAPTTTKEPAPTTTKEPAPTT 724
Db 7438 EVPEKVEKVPKPAVAVKKEVPKAPVEVPKVP-----LEKPAVP--VPEAPEBPPEV 7490
Qy 725 PKKPAKELAPTTTKEPSTTSDKAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 784
Db 7491 YEB--PEEIAF--EEEIAPEEKVPVAEE--EEBVPVPAPEBPBKIIPEKVP- 7540
Qy 785 TLKEPAPTTPKKPAKELAPTTTKEPSTTSDKAPTTTKEPAPTTTKEPAPTTTKEP 837
Db 7541 VIKKEAPPPKPEBEKVP--IEKPKLRPPPPPPAPKEDVKEKIFQLKAIPEKVP 7597
Qy 838 TPKEP-APTPEPTPTTSEVS--TPTTKEP-----TTIHSPEESTPELSAETP 886
Db 7598 NPQVEKVELTPKVPGEKVKRKLPERKEPEKEVVLKSVLRKPEEPEKVE---P 7653
Qy 887 KALENSPKPEGVPTTTPAATKP-----EMTTAKDTERDLATPETTTAAPKTETA 942
Db 7654 KKLK-KVKKAPV--EPPPEKVEVEVPTTTERKRIPEPTKPELEKPAIPLAPBP 7709
Qy 943 TTTKETSKTATTTQVSTTTQDTPPKITTLKTTTLAPKVTTKTITTTIEMK- 1000
Db 7710 PKPEA-----EVTIKPPEVEPEPTIAAPVVPVAVGKAAE 7745
Qy 1001 -----PEETAKPKDARTNSKATTPKQKPTKAPKKTTS--TKKPKMPVRKPKTTP 1052
Db 7746 AKAPKEEAKPPGPIKVPKKTSPSILEARRKLARGSGGKEXPDEAPPTTQKAVLKV 7805
Qy 1053 ---RMTSTPELNPSTRIAEAMLOTTT 1077
Db 7806 KEIKDIIITSEFVGSSAIFECVSPST 7833

```

RESULT 12

745025

```

mucin MUC5B, tracheobronchial [imported] - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jul-2000
C:Accession: U45025
R:Desseyn, J.L.; Guyonnet-Duperrat, V.; Porchet, N.; Aubert, J.P.; Iaine, A.
J. Biol. Chem. 272, 3168-3178, 1997
A>Title: Human mucin gene MUC5B, the 10.7 kb large central exon encodes various alternate
A:Reference number: 228999; MUID:97166151; PMID:9013550
A:Accession: U45025
A>Status: preliminary; translated from GE/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-3570 <DES>
A:Cross-references: EMBL:Z72496; NID:G1834502; PIDN:CAA96577.1; PID:G1834503
A:Experimental source: placenta
C:Genetics:
A:Gene: MUC5B

Query Match. 8.8%; Score 663; DB 2; Length 3570;
Beat Local Similarity 22.1%; Pred. No. 8.7e-20;
Matches 396; Conservative 119; Mismatches 578; Indels 698; Gaps 72;

Qy 54 MECCPDPRKVCVTAALSCKGRCFESFERGECDCDAQCKDYDCPYE--SFCAEV-HNP 110
Db 1581 VECSLDGLGVC-----NRR-----QVGF-KMCFNYEIRVFCNNGHCP 1619
Qy 111 TSPSSKAPPS-----GASQIKSTTKRSPKPKKTKKYIE 150
Db 1620 STPATSTPATSTSPGTGTTWILTEQTAATTATTGTAISSTGTGAPP-----KVL 1673
Qy 151 SEETIEHVSSENOESS-----SSSSSSSSSTIKWS---KNSANR--- 192
Db 1674 SQATTPTATSKATSSSPRATTLPLVLTSTATSTATSPFIPISSTLGTGTSQNNP 1733
Qy 193 -----ELQKKLVKONKKNRTKKKTPRP-----VDE- 224
Db 1734 PMATMSTIHPSSTPEHTHTSTVLTTKATTTATSTSMSTSPSTGTTWILTELTAA 1793
Qy 225 GLDNG-----DEKVTTPDSTQNH-----KVSTPKITTKAPI 258
Db 1794 ALPHGTSPSTPGTTWILTEBSTTATVTPVPGSTATASTATAGTAKVLIS-TATTP 1852
Qy 259 NFR-----PSLPNSDTSKETSLTVNKEVET-----KETTTNKQTSIDG 300
Db 1853 SSRATPSSPGTATALPALNSTATTPTATSTVTAIPSSLSGTAWRLSQTTPATM 1912
Qy 301 KEKTSAKETOSIKTSKADLPATSKVLAKPTPAETTK-----GPA 343
Db 1913 PSTPEVHTSTVLTATTAITRTGSVAPSPSTPGTAHTTKVPTTTTGFTATPSS 1972
Qy 344 LTPP-----KEPTPT-----TPKE-----PASTPKPEPTP 368
Db 1973 LTPPVWISTTTTPTTSGSTVTPSSI PGTHATVLTITTTTVAATGSMATPSSS 2032
Qy 369 TTIKAPTTKEPAPTTTTSAP--TPKEPAPTT-KEPAPT-----TPKEPAPTTKE 421
Db 2093 PSLTATTAATTANSTNNPSTPGTPIPLVLTATTAATPAATSSVPSALGTHHPV 2092
Qy 422 PTTTKSA-----PTTP-----KEPAPTTPKKAPPTT--KEP 451
Db 2093 PNTATTHGRSLPPSSSHVPTAWTSATSGILGTHTEBGTGSHTPAATGTGTQSTP 2152
Qy 452 APTTPEPTPTTKEPAPTTKEPAPT--TPKEPAPAKKAPPTTKEPAPTT--PKEPAP 508
Db 2153 ALSSPHSSRTTSEPPSGTTTPGHTGTSKRTTATATPSKTRISTLLPSSFTSAP 2212
Qy 509 TTKPEPS----- 515
Db 2213 TTGCEPQCAEMSEWLDYVPMGPBGDPDTYSNIRAAGAVCEQPLGECQAQAPG 2272
Qy 516 -----PTTTPKEPAPTTTTSAP 531
Db 2273 RELQGVVECSLDGLGVRNREQVGKFMCFNYEIRVFCNNGHCPSTPATSTPAT 2332

```

[illegible]

C.Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C.Accession: T34513
R.Favella, A.; Vaudin, M.
submitted to the EMBL Data Library, August 1994
A.Description: The sequence of C. elegans cosmid ZK783.
A.Reference number: Z21536
A.Accession: T34513
A.Stratus: preliminary; translated from GB/EMBL/DBJ
A.Molecule type: DNA
A.Residues: 1-3507 <EAV>
A.Cross-references: EMBL:U13646; PIDN:AAC24418.1; GSPDB:GNO0021; CESP:ZK783.1
A.Experimental source: strain Bristol N2; clone ZK783
C.Genetics:
A.Gene: CESP:ZK783.1
A.Map position: 3
A.Introns: 14/1; 48/1; 84/1; 196/3; 303/1; 381/1; 586/1; 605/1; 1175/3; 1207/1; 1409/2;
3504/1

```

Query March 8.7%: Score 658.5; DB 2; Length 3507;
Best Local Similarity 22.6%: Pred. No.1.3e-19;
Matched 286; Conservative 178; Mismatches 484; Indels 317; Gaps 51;

QY 27 LSSAGRCGSGEYSKD-ATC-----NCDVNCOHMECCPDKFRKCTAELSCGRCFESF 78
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1781 MGSGCGCKMACYTGDGATCIIEEPEKDKTA-----CTDEMSRICELE---KKQCTYDE 1832
QY 79 ERGRBC-----DCAQCKKYDKCCPDYBSFCAEVHNPTSPSSKKAP 121
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1833 EEVPOCGACLPQHHPINGTCSLQISGLCAQKNCNKAH--CIDIH-----PDSHFQSC 1885
QY 122 PSG-----AQTIKSTTKRPRKPNKKTKVISEEIT 155
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1886 PDGPIGDMICDDYDECCNAGCCDENTKCENTIGSFNVCCLGEGKKYKDEKCVVDEKOP 1945
QY 156 EEHSVSENOESSSSSSSSSSSTIMKIKSSKNSANRELQKTLKVONKKURTAKKPPPK 215
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1946 NREKIEIDENSSSNSG-----GKPPTK 1970
QY 216 PPVYDEAGSLDNGDFKTTPTDTSTQHNKSTSPKITTAKPINRPLPRP---SDRSK 272
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1971 GIVASTS-----ATSEESTAPPHATTSSISTTSKDMTSSKSPENVTVSSSP 2019
QY 273 ETSLVNKEETV-ETKEETTNKOTSDGKEKTSIAKETOSIEKTSADLAP---TSKV 327
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2020 EVSTSSSSKSTTASSETYVSTSPSSSSSSSEAPLTSEATTBVTBSSSVKSTPKESSESI 2079
QY 328 LAKPTPKAETTKGPALTTPKEPTPTTPKEBASTTPKEPTPTIKS-APTTPKEBAPTTT 386
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2080 TVKLSSKSPETBESSVKSSPSTPS-TTSQSVTSIVPETSKSTVSSSEAPVSTSPTEVHT 2138
QY 387 KSAPTTPKEPAPTTTTPKEBAPTTTPKEBAPTTTTPKAPTTTTPKAPPTPKKAPPT 446
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2139 -SSETKSLSSASTTGDNSTTPTSTSLASVKSASAPGTSASVAPVLSLSPVSSOPS 2197
QY 447 TPK-EPAPTTPKEPPTTPTPKEBAPTTKEP-APTTPKEBAPPAKPPAPTTTPKEBAPTPK 504
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2198 TKTFDARESSVVASSEHSSTSGTSVKSSTSEPHVTKLSTSSNPSSSVAVTSPKSPPTYPE 2257
QY 505 EPAPTTTKESPPTP--KEPAPTTKSAPTTPKEBAPTTKSA-----TTPKEPSPTT 556
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2258 -----STEQPSTTSPGQSLTPMNSNSEVLTTSEPHVLLSSLPVSSQSGSTTPNNLSSESS 2312
QY 557 TKEBAPTTTPKEBAPTTTPKKEBAPTTTPKEBAPTTTPKEBAPTT----- 596
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2313 TVE-----TPKSSSEVLSNSEPSTTEAFTLSPDILSTTNNSQSGSTVSTEDRSEISE 2368
QY 597 -TKKPAPLAPK-----EPAPTTPEKAPTTPKKLATTPTEKLAATTPPEKAPPTP 645
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2369 NSEKPT-SAPELVNSVTVHVASSPDVPTES--SEPDOLTSSTENIPEBASQKQISPT 2425
QY 646 BELAPTTPEEPTPTTPPEP-----APTTPKAAAPPTTPKEBAPTT----- 684
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2426 TPDVTTASSEETKSTSMSPDLSTSNVLSSESTTPSESS---KSPVSSSTEGISVVTST 2481

```

QY 685 --PKKAPPT-----TPKEPAP-----TPKKEPAPPTKGTAPPTLKKEPAPPTPKK 728
DB 2482 EFSKVPESSTISSVLBEDLTKTTPSPILKEETTTASSETSEPLETDSLTVSRHIELTSSBN 2541
QY 729 APKELAPPTTKKPTSTSDKPA-----PT-----TP-----KGTAP- 759
DB 2542 VPKESSTTSSESSEKPSGEPAGILTSVTVVPTSSVSLTASEIAITSNPFKGRPI 2601
QY 760 -TTPEK-----EPAPPTKKEPAPPTKGTAPPTLKKEPAPPTPKKAPKELAPPTTKGPT 811
DB 2602 TTPSPSLVKSSTTSPSTVTSSEPSSESTKRTVSTVSTTPTTEETTSSESLITTAAPSKPT 2661
QY 812 -STSDKAPPTTKKAPPTTKKEPAPPTPKKAPPTTPETPTPTSEVSTPTTKAPTIN 870
DB 2662 ESTTSSAPPTTKAPPTTKSETKPSNVSTSKSTENVTSTSGSLSSSTMSST- 2715
QY 871 KSPDESTP--ELSAEPTPKALE--NSPKBPVPTTKPAATK-----PEMTTAKDKTTE 921
DB 2716 SEPETNAPAVTVSSASSTLEENSTSSP-----TSSEASVLSLFPESITSEAVTVSS 2771
QY 922 RDLRTPTTTTAPKMTETATTTKTESKITAATT-----QVTSTTQ 966
DB 2772 R-----APAEITWSSSHREISTVSESPSEPEIPILSTVSPNVVTASSIPSEEPILSVTS 2827
QY 967 DTPP-FKITT-----LKTTLAPKVTTTKITTEIMNKPEETAKPKDRATNSKATPK 1020
DB 2828 STPTVRAITGTPDDLIVSVTVPSHGRRONITASSV--PSNPSPIILPSES-LITPQ 2883
QY 1021 POKPTAKPKPTSTKPKPTMPRVKPK--TTPTRKMTSTWPELN-----PTSRIBAM 1072
DB 2884 PPTTTTAKAPATTSCKKGPPEIOPPAEMFTTPAP-----PPENGVGGEETNGEEOV 2937
QY 1073 LQTTT 1077
DB 2938 TSTTT 2942

RESULT 14

151618
nucleolar phosphoprotein - African clawed frog
C:Species: Xenopus laevis (African clawed frog)
C>Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 24-Sep-1999
C:Accession: I51618; #557757
R:Calins, C.; McStay, B
J:Cell Sci. 108, 3339-3347, 1995
A:Title: Identification and cDNA cloning of a Xenopus nucleolar phosphoprotein, xNopp180
A:Reference number: I51618; MUID:96019267; PMID:7593294
A:Accession: I51618
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-990 <CAI>
A:Cross-references: EMBL:X88927; NID:g895920; PIDN:CAA61368.1; PID:g895921
C:Gene: xNopp180
C:Superfamily: nucleolar-cytoplasm shuttle phosphoprotein
C:Keywords: phosphoprotein

Query Match 8.6%; Score 643.5; DB 2; Length 990;
Best Local Similarity 24.2%; Pred. No. 1.6e-19;
Matches 246; Conservative 148; Mismatches 424; Indels 159; Gaps 40;

QY 112 SPPSKKAPPPSG-----ASQTKSTTKSRKPKPKKT-----KYIESREITEE 157
DB 57 SPDACKRRPPANGLPKKKSANESSSEDSSEEDPPAKKAQPAAGKPPVKAQVPPKA 116
QY 158 HVSVENQSSSSSSSSSSSTIWK--IKSKNSAARELOKCLKVNDKNKNT-----KK 210
DB 117 KSSSEDSDESDSEETTKKPPAKRAQTPKAAVVTPTQKAKASSSSSSSEDSASKK 176
QY 211 KPTPKP--VVDKAGSLDNGDFVTTPTDSTTGHNVSTSPKITTAKPINRPSLPN 267
DB 177 QPVIKVPKQAVVAKGLASNG--KTADSSSSSDSDSPAKKTAATK-----TTP- 224

QY 268 SDTSKETSIVNKEETVETKETTNTKQTSNKGKTTSAKETOSIEKTSADLAPTS-- 325
DB 225 --TPRATIAKQAKTKAGKSSSSREDSDESDSEOKTAKSKAPKDVSAV--PPTSVS 279
QY 326 --KVLAKPTPAETTTKGPALTTTPKEPTTPKE--PASTTPKEPTPTTIKAPPTPKP 381
DB 280 KKTLLSQGTAKAESSSSDSDEBQPAKKAVIPAKAASAPFLAKAEITSGES 339
QY 382 APPT-----TSAPPTKKEPAPPTTKKEPAPPTTKKE--PAPPTTSAP 429
DB 340 DSSSEDEKSSVKLGVNAAP--KKAFA-----APAKSTPVAAKKSAPAKASSSSD 390
QY 430 TTPKEPAPTTKDPAPTTKKEPAPTTTPKEPTTPKKEPAPTTKKEPAPTPK 488
DB 391 SDSSNEETTTKPAKATTTPAKSA--ATPSTKPTNGK--ATPSTKPAKAGTPTKSTAKDS 448
QY 489 KPAETTPKEPAPTTKKEPAPTTKPEPTTPKEPAPTTKSAPTTKEPAPTTKSAPTT 548
DB 449 SSSDSSDSSEETTTKPAKATTTPAKSAATPTSKPTNSKATPTSKTPAKGTPKTSAA 508
QY 549 PKESPTTKKEPAPTTKKEPAPTTKKEPAPTTKKEP--APTTPKEPAPTTTKKAPTAKE 607
DB 509 KKDSSSDSSSDSDEKTPA--KRAKTTKPAKPAKTTAKPA-----AKTTPAK 557
QY 608 PAPTPKETAFTTKKLTPTTPEKLAPTTPEKAPAPTTPELAPTTPEEPTTPEEAPPT 667
DB 558 PA-----AKSTPEKQVPTKE-----SSSDSSSSSEDEKKS 590
QY 668 TPKAAPTPKEPAPTTKKEPAPTTKKEPAPTTKETAFTTKGTAFTTLKEPAPTTPK 727
DB 591 SAKAVKTT--PKATSK--PVVASKPV--AKKASSSDSDESEETTKTTPPLTGLS 643
QY 728 PAPKELAPTTKERTSTSD-----KPAETTKGTAFTTKKPAAPTTKEPAPTTPKG 780
DB 644 PAVKTLPEKKAESSSDSDESEKTKPAKPAKSAATPVNTKAPAONKASKASCSDS 703
QY 781 TAPTLKEPAPTTKPKAPKELAPTTKGPSTTSDPAPTTKETAFTTKKEPAPTTPK 840
DB 704 SSEEGKSKOP--TKSPAKATAP--PKNPVAVNKDPSSSSSSSDSDEKQ--KPK 757
QY 841 KPAETTPETPTTSEVSTPTTKEPTTIHKSPESTPELSAEPKALENSPEKPGVPT 900
DB 758 QAA-----AAKOVQOAKAKKPTPKKAASSSE--DS 787
QY 901 TKTPAATPEMTTAKOKTTERDLRTTPETTTAPKATKEATTTKTESKITAATTQV 960
DB 788 SSDESVKAKKTNVAVSKSPV--TTPKAVPAKKSSESSSEDEKQGGKNTSTTKI 843
QY 961 TSTTQDTPPKITLTKTTTLAPKVTTTK--KITTEIMNKPEETA--KPKDRATNS-- 1014
DB 844 ANST-----PKAAAACSESSSEDEKANGTSGKRGESTGNAEC 885
QY 1015 KATTPKQKPTAKKPKPTSTKPKPTMPRVKPKTTPP--RKMTSTWPELNPTSRIAE 1070
DB 886 EAVTPE-----NKGLKAKSPNTFPKVKKEKLNTPFRVVEDIEINP--RMAD 932

RESULT 15

711622
extensin class 1 precursor - cowpea
C:Species: Vigna unguiculata (cowpea)
C>Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 17-Nov-2000
C:Accession: T11622; S54155
R:Arseanjlevic-Maksimovic, I.; Broughton, W.J.; Krause, A.
Mol. Plant Microbe Interact. 10, 95-101, 1997
A:Title: Rhizobia modulate root-hair-specific expression of extensin genes.
A:Reference number: Z17301; MUID:97155574; PMID:9002273
A:Accession: T11622
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-489 <ARS>
A:Cross-references: EMBL:X91836; NID:g1015936; PIDN:CAA62943.1; PID:g1015937

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 8, 2003, 09:20:36 ; Search time 18 seconds

(without alignment)
3668.083 Million cell updates/sec

Title: US-09-556-246-1

Perfect score: 7526

Sequence: 1 MAMKTLPIYLTLISVFIYQ.....AAATTRSGQTLKVMWNC 1404

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1188	15.8	5179	1 MUC2_HUMAN	Q02817 homo sapien
2	950	12.6	1664	1 SLPI_CLOTM	Q06852 clostridium
3	809	10.7	1367	1 AMYH_YEAST	P08640 saccharomyc
4	707.5	9.4	555	1 GPI_CHLRE	Q9F866 chlamydomon
5	703.5	9.3	5085	1 PCLO_RAT	Q9JF86 rattus norv
6	665	8.8	5038	1 PCLO_MOUSE	Q9GTX7 mus musculu
7	660	8.8	5703	1 MUSB_HUMAN	Q9HCV0 homo sapien
8	659.5	8.8	5147	1 PCLO_HUMAN	P15941 h mucin 1 p
9	651	8.7	1255	1 MUC1_HUMAN	Q01761 caenorhabdi
10	629	8.4	6632	1 UNB9_CAEBL	Q9PU36 gallus gall
11	578.5	7.7	5120	1 PCLO_CHICK	Q9PU33 homo sapien
12	573	7.6	2812	1 ZAN_HUMAN	Q9Y493 homo sapien
13	559	7.4	875	1 FPI_MYTEB	Q25460 mytilus edu
14	551	7.3	620	1 EXTN_TOBAC	P13983 nicotiana t
15	551	7.3	1087	1 NPH_MOUSE	P19246 mus musculu
16	530.5	7.0	1162	1 TCNA_TRYCR	P23253 trypanosoma
17	530	7.0	865	1 CPN_DROME	Q02910 drosophila
18	518.5	6.9	872	1 FPI_MYTGO	Q25434 mytilus
19	517.5	6.9	831	1 NPI_RAT	Q16884 rattus norv
20	509.5	6.8	662	1 MUC1_XENLA	Q05049 xenopus lae
21	503.5	6.7	1970	1 RPB1_HUMAN	P24928 homo sapien
22	502	6.6	1020	1 NPH_HUMAN	P12365 homo sapien
23	498.5	6.6	1970	1 RPB1_MOUSE	P08775 mus musculu
24	493.5	6.6	467	1 RPB1_CRIGR	P11414 cricetus
25	490	6.5	2142	1 BAT2_HUMAN	Q04834 homo sapien
26	488.5	6.5	826	1 SSP2_PLAYO	Q01443 plasmodium
27	478.5	6.4	857	1 APPI_SCHPO	Q9P788 schizosacch
28	475.5	6.3	267	1 EXTN_MAZE	P14918 zea mays (m
29	471.5	6.3	5376	1 ZAN_MOUSE	Q08799 mus musculu
30	469.5	6.2	1509	1 GSRI_HUMAN	Q9H2M4 homo sapien
31	468.5	6.2	634	1 HWP1_CANAL	P46593 candida alb
32	467	6.2	817	1 VRP1_YEAST	P37370 saccharomyc
33	454	6.0	797	1 VGLX_HSVB	P28968 equine herp

34	454	6.0	1161	1 DAN4_YEAST	P47179 saccharomyc
35	453.5	6.0	1794	1 YDC9_SCHPO	Q10172 schizosacch
36	450	6.0	3178	1 YB89_CAEBL	Q09692 caenorhabdi
37	448.5	6.0	670	1 VGS0_HSV1	Q00130 iccaturid h
38	446	5.9	1229	1 P121_HUMAN	Q9Y2N1 homo sapien
39	444.5	5.9	2774	1 MABA_RAT	P34926 rattus norv
40	442.5	5.9	751	1 FPI_MYTGA	Q27409 mytilus gal
41	439.5	5.8	1083	1 T2D3_HUMAN	Q00268 homo sapien
42	436.5	5.8	1411	1 TCOP_HUMAN	Q14428 homo sapien
43	434.5	5.8	3256	1 KIGU_HUMAN	P46013 homo sapien
44	433.5	5.8	3164	1 TEGU_HSV1	P10220 herpes simp
45	432.5	5.7	439	1 XP2_XENLA	P17437 xenopus lae

ALIGNMENTS

RESULT 1
MUC2_HUMAN STANDARD; PRT; 5179 AA.
AC 002817; Q14878; 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 41, Last annotation update)
DE 28-FEB-2003 (Rel. 41, Last annotation update)
DE Mucin 2 precursor (intestinal mucin 2).
GN MUC2 OR SMUC.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Intestine;
RX MEDLINE=94132002; PubMed=8300571;
RA Gum J.R., Jr., Hicks J.W., Toribara N.W., Siddiki B., Kim Y.S.;
RT Identification of the amino terminus and overall sequence similarity
RT to prepro-von Willebrand factor";
RL J. Biol. Chem. 269:2440-2446(1994).
RN [2]
RP SEQUENCE OF 626-1895 AND 4196-5179 FROM N.A.
RC TISSUE=Colon;
RX MEDLINE=93016075; PubMed=1400449;
RA Gum J.R., Jr., Hicks J.W., Toribara N.W., Roche E.-M., Lagace R.B.,
RT Kim Y.S.;
RT "The human MUC2 intestinal mucin has cysteine-rich subdomains located
RT both upstream and downstream of its central repetitive region.";
RL J. Biol. Chem. 267:21375-21383(1992).
RN [3]
RP SEQUENCE OF 1343-1895 AND 4176-4195 FROM N.A.
RX MEDLINE=9358717; PubMed=1885763;
RA Toribara N.W., Gum J.R., Jr., Culhane P.J., Lagace R.E., Hicks J.W.,
RT Petersen G.M., Kim Y.S.;
RT "MUC-2 human small intestinal mucin gene structure. Repeated arrays
RT and polymorphism.";
RL J. Clin. Invest. 88:1005-1013(1991).
CC - FUNCTION: COATS THE EPITHELIA OF THE INTESTINES, AIRWAYS, AND
CC OTHER MUCOS MEMBRANE-CONTAINING ORGANS. THOUGHT TO PROVIDE A
CC PROTECTIVE, LUBRICATING BARRIER AGAINST PARTICLES AND INFECTIOUS
CC AGENTS AT MUCOSAL SURFACES.
CC - SUBUNIT: MULTIMERIC.
CC - TISSUE SPECIFICITY: COLON, SMALL INTESTINE, COLONIC TUMORS,
CC BRONCHUS, CERVIX AND GALL BLADDER.
CC - PTM: ALL CYSTEINE RESIDUES ARE INVOLVED IN INTRACHAIN OR
CC INTERCHAIN DISULFIDE BONDS (BY SIMILARITY).
CC - POLYMORPHISM: THE NUMBER OF REPEATS IS HIGHLY POLYMORPHIC AND
CC VARIES AMONG DIFFERENT ALLELES.
CC - SIMILARITY: Contains 1 C-terminal cysteine knoc-1like (CTCK) domain.
CC - SIMILARITY: Contains 1 TIL (trypsin inhibitory-like) domain.
CC - SIMILARITY: Contains 2 WWFC domains.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration


```

QY 546 P---TTPEKESPTTKEBAPPTTKEBAPPTPKKAPPTTKEBAPPTTKEBAPPTTKEBAP 602
DB 1699 PSEPTTSSPI-TTTTSSSTTSPSPPTTWTTPSPPTTSPPTTMTTLTPPTTSSPLT 1757
QY 603 TAPKEAPPTP-----KETAPPTP----- 621
DB 1758 TTPPLPSITPPTSPSTTPTTTPCVPLCNMTGMLDSGKFNKPGDTELLGDVCGPGW 1817
QY 622 -----KKLTP----- 626
DB 1818 ANNISGRATMPVPVPIGQLGQTVICVSVGLCKNEDQKGVIPAFCLNTEINVOCEB 1877
QY 627 --TTPKELAPTTPEKAPPTTPEELAPTTPEEPPTT--TPEBAPPTPKAAAPNTPEKAPPT 683
DB 1878 CVTQPTTMTTTEENTPPTTPTTITTTTTPPTPTGQTPTTITTTTITTTTTPPTPT 1937
QY 684 TPEKP-----APTPEKAPPT--TPEKAPPTPKGTAFTTLKEBAPPTPKKAPKELAPPT 737
DB 1938 GTQPTPTTPTTPTTPTTPTPTPTGQTPTTPTTPTTPTTPTTPTPTGQTPTTPTTPTT 1997
QY 738 TKEPTSTSDKAP--TTPKGTAPTTPEKAPPTTPEKAPPTPKGTAFTTLKEBAPPTPKK 796
DB 1998 TTPPTPTPTGQTPTTPTTPTTPTTPTTPTPTPTGQTPTT--TTPITTTTTPPTPTPTGQT 2056
QY 797 PAKELAPTTTKEGPTSTSDKAP--TTPKETAFTTPEKAPPTTPKKAPPTTPEPTPTTS 855
DB 2057 PTTTPTTPTTPTTPTPTPTGQTPTTPTTPTTPTTPTTPTPTPTPTPTPTPTPTPTPTT 2115
QY 856 EVSTPTTKEPTTIHKSPESTPELSAETPKALENSPEK-----GVPT--TKTP 904
DB 2116 PTPPTPTGQTPTT---TPTTPTTPTTPTPTPTGQTPTTPTTPTTPTTPTPTPTPTPT 2172
QY 905 AATPEMTTAKOTERBDR---TTP--ETTAAKMT---KETATTEKTESKATPT 956
DB 2173 TTPPTTPTTPTTPTPTPTPTGQTPTTPTTPTTPTTPTTPTPTPTPTPTPTPTPTPTPT 2232
QY 957 TTPVSTTTPQDTPPTPKITTLKTTTLAPKVTTL--KTTITTEIMNKDEETAKPDORATNSK 1015
DB 2233 PTP--TGTPTPTTP--ITT--TTTTPPTPTPTGQTPTTPTTPTTPTTPTPTPTPTPTPT 2287
QY 1016 ATTPKQKPTKAP--KKTSTKPKTMPRVAKKPTTTPPKMTSTMBELNP--TSRIABA 1071
DB 2288 TTPPTTPTTPTTPTPTPTPTGQTPTTPTTPTTPTTPTTPTPTPTPTPTPTPTPTPTPT 2347
QY 1072 MGTTPR--TNGPNSKLVENPKSEBAGAEGETPHMLARPHVMEBTP 1120
DB 2348 PTPPTGQTPTTPTTPTTPTTPTPTPTPTG--TQTP--TTPPTTPTTPTTPT 2392

```

```

CC -1- SUBCELLULAR LOCATION: Cell wall.
CC -1- SIMILARITY: Contains 4 S-layer homology (SLH) domains.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.1sb-sib.ch/announce/
CC or send an email to license@sib.ch).
CC -----
CC DR EMBL; X67506; CAA7841.1; -.
CC DR PIR; T18262; T18262.
CC DR InterPro; IPR001119; SLH.
CC DR Pfam; PF00395; SLH; 3.
CC DR PROSITE; PS01072; SLH_DOMAIN; 2.
CC KW Cell wall; S-layer; Signal; Repeat.
CC FT SIGNAL 1 28
CC FT CHAIN 29 1664
CC FT DOMAIN 36 763
CC FT REPEAT 36 191
CC FT REPEAT 207 363
CC FT REPEAT 409 565
CC FT REPEAT 607 763
CC FT DOMAIN 771 1377
CC FT FT
CC FT DOMAIN 1378 1449
CC FT DOMAIN 1453 1494
CC FT DOMAIN 1495 1565
CC FT DOMAIN 1566 1625
CC FT DOMAIN 1626 1646
CC FT DOMAIN 1664 178194 MM; 5F39695BBAFE74B CRC64;
CC SQ SEQUENCE 1664 AA; 178194 MM; 5F39695BBAFE74B CRC64;

Query Match 12.6%; Score 950; DB 1; Length 1664;
Best Local Similarity 31.0%; Pred. No. 2; 9e-29;
Matches 312; Conservative 104; Mismatches 361; Indels 230; Gaps 55;

QY 327 VLAKPTP--KAETTTGPAALTTPKEPTPTTTPKEBAPTTTTPKEBAPTTTTPKEBAPTT 385
DB 758 VVIQPAPIKASDEBIPDTSPDEPTPS-----DEPTPS--DEPTPSDEPTSD 804
QY 386 TKSAPTPEKAPPTTTPKEBAPTTTPKEBA---PTTTPKEBAPTTTTSAPTTPKEBAPTTPK 441
DB 805 EPTPSBETPEBIPDTSPDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSBETPEBIPDTSP 864
QY 442 KPAFTTPKEBAPTTTTPKEPTTTPTPKEBAPTTTPKEBAPTT--TPKEBAPTAKKAPPTTPEKAP 500
DB 865 SDEPTPSDEPTPS--DEPTPS--DEPTPS--SDEPTPSBETPEBIPDTSPDEPTPSDEPTP 919
QY 501 TTPKEBA--PTTTPKEBPTT--TPKEBAPTTTTS--APTTPKEBAPTT--TTSKAPTTPKES 553
DB 920 SDEPTPSBETPEBIPDTSPDEPTPSBETPEBIPDTSPDEPTPSBETPEBIPDTSPDEPT 979
QY 554 ----PTTTPKEBAPT--TPKEBAPTTTPKAPPTTPKEBAPTTTPKEBAPTTTTPKAPTAPEK 608
DB 980 PSDEPTPSBETPEBIPDTSPDEPTPSBETPEBIPDTSPDEPTPSDEPTPSDEPTPSDEPT 1039
QY 609 APT---TPKETAPTTPKLPTTPPEKLAPTTPEKAPTT--TPEBAPTTTPEBPTT--TP 661
DB 1040 TSDSDEPTSEB-----TPEBIPDTSPDEPTPSDEPTPSDEPTPSDEPTPSDEPTSETP 1094
QY 662 EEPATTTPKAAAPNTPEKAPPTTTPKEBA--PTTTPKEBAPT--TPKETAPTTTTPGTAFTTLK 718
DB 1095 EEPITPTSPDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSBETPEBIPDTSPDEPTSD 1154
QY 719 EAPTTTPKKAPKELAPTTTKEPT--STSDKAPPTTTPGTAFTTTPKEBAPTTTPKEBAPTT 777
DB 1155 EPTPS--DEPTPSD--EPTPSBETPEBIPDTSPDEPTPSDEPTPSDEPTPS--DEPTPS- 1208
QY 778 PKGTAFTTLKEBAPT--TPKKAPKELAPTTTTPGTAFTTSTSDKAPTT--TPKETAPTTPKE 833
DB 1209 ---DEPTPSDEPTPSBETPEBIP-----PDTSPDEPTPSDEPTPSD--EPTPSDE 1253

```


Gy		605	PKEBAPPTPKKPAATTPKKLAFPTTECKALAPTPEKAPATTPEELAPLT--TPEPPTTPEE	663
Dd		612	PPSSSTTBESSAPVT-----SSTTBSSAP-VPTSSSTTBSSSAVPVPTSSSTTBSSA	666
Gy		664	PAP-----TTPKAAPNTPEKBPATTPKEBAPTTPKEBAPTTPKETAPT-TPKGAPPTL	717
Dd		667	PVPPTSSSTTBESSAPVT-----SSTTBSSAPVT-----SSTTBSSAVPVPTSSSTTBSS	718
Gy		718	KEPATTPPKKAPKLAPTTTKKEPTSTTDKAPPTPKGIAPTYTKEAP---TTPKPA	774
Dd		719	SAPVPTPSSTTBESSAPVPTPS-SSTTBSSAPVTSSTTBSSAPVPTPSSTTBSSA	777
Gy		775	PT-TPKGAPTLKBPATTPPKKAPKBLAPTTGPISTSDKAPPTPKETAATPKXE	833
Dd		778	PVPPTSSSTTBESSAPVPTPSSTTBSSAPVPTPSSSNITTSAPSSPTPSSTTBSSV	837
Gy		834	PAP-----TTPKBPATTPETPTPTSEVSFT-----TKKEPTIHSPDSTPELSAE	883
Dd		838	PVPPTSSSTTBESSAPVSSSTTBSSAVAPVPTPSSSNITTSAPSI---PSSSTTBEST	894
Gy		884	PTPKALENSPKBPGVPTTKPAIKPEMTTAKDKTTERDRTKTPETTPAAPKMKTETAT	943
Dd		895	GT-TVTPSSSKPKSGQTESVASTTETTIVPTKTTSVTPTSTITTITTVCGSTGNAGE	953
Gy		944	TTEKTEBKIRAT--TTOVSTSTTODTTPRKTIILKTTLAAPKYT---TKKITITT-EI	997
Dd		954	TTCGSPKTVTTTTVPTTTTTSVTSSTTTITTVCGSTGNAGETSCGSKRITTTVPC	1013
Gy		998	MNKPEETAKPKDRATNSKATTPKPQP-----TKAPKKEPTSTKPKXTM	1040
Dd		1014	STSPEBA-----SETTISPTTPTVVSTTVTTEYSTSKPGELITTTFTVKNI	1055
Gy		1041	PRVRKPKTTPPRKMTSTMPELNPTSRIAEAMLQTTTRNOTPNKSCLVENPKSEDAGA	1100
Dd		1066	PTYTLTTIATPT--SVTTVTNFTPTT-LTTVCSTGT-----NSAGBTSGC	1109
Gy		1101	EGERPHMLLRHVFMPEVTTPMDYLPKRPNGCIINPMLSDETINCCKPVDGLTTLANG	1160
Dd		1110	SPKVTTTTVPCSTGGEYTEETLVTTAVTTVVTTTBSSSYGTMSA-GKITTGYYTMSVP	1168
Gy		1161	TLVAFRGHYFWMLSPPSPSPARITEVMGISPIDVTFTRC---NCEGKT	1208
Dd		1169	TT-----YTTLAPSAFVTAINT-----AVPTTITT--TECSAATNAGET	1207

RESULT 4
GPI_CHURE STANDARD; PRT; 555 AA.

AC	OQFFQG;	003927;	
DT	28-FEB-2003	(Rel. 41,	Created)
DT	28-FEB-2003	(Rel. 41,	Last sequence update)
DT	28-FEB-2003	(Rel. 41,	Last annotation update)
DE	Vegetative cell wall protein gpi precursor (Hydroxyproline-rich glycoprotein 1).		
OS	Gpi.		
ON	Chlamydomonas reinhardtii.		
OC	Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;		
CC	Chlamydomonadales; Chlamydomonadaceae; Chlamydomonadaceae.		
CX	NCB1	_Taxid=3055;	
RN			
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=21159092;	PubMed=11258910;	
RA	Fertile P.J., Woeseener J.P., Waffenschmidt S., Kilz S., Drees J.,		
RA	Goodenough U.W.,		
RT	"Glycosylated polypurine II rods-with-links as a structural motif in plant hydroxyproline-rich glycoproteins.";		
RL	Biochemistry 40:2978-2987(2001).		
RN	[2]		
RP	PARTIAL PRELIMINARY SEQUENCE FROM N.A.		
RX	MEDLINE=91017504;	PubMed=1699225;	
RA	Adair W.S., Apt K.E.;		
RT	"Cell wall regeneration in Chlamydomonas: accumulation of mRNAs encoding cell wall hydroxyproline-rich glycoproteins.";		

```

BL Proc. Natl. Acad. Sci. U.S.A. 87:7355-7359(1990).
CC -I- FUNCTION: Major component of the outer cell wall w6 (crystalline
CC layer.
CC -I- SUBUNIT: Associates with gp2 and gp3.
CC -I- PTM: N-glycosylated and O-glycosylated.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements
CC or send an email to license@ebi.ac.uk).
CC -----
DR EMBL; AF309494; AAC645420.1; -.
DR EMBL; M58496; AAA69706.1; ALT_SEQ.
DR GLYCOSUB; O9P06; -.
DR InterPro; IPR002965; P_rich_extensin.
DR InterPro; IPR003882; PlstII_extensin.
DR PRINTS; PRO1217; PRICHEXTENS.
DR PRINTS; PRO1218; PSTLEXTENSIN.
KW Glycoprotein; Repeat; Signal.
FT SIGNAL 1 29
FT CHAIN 30 555
FT DOMAIN 40 339
FT DOMAIN 259 279
FT CARBOHYD 399 399
FT CARBOHYD 455 455
FT CARBOHYD 493 493
FT SEQUENCE 555 AA; 54219 MW; 6A584A90465502F5 CRC64;
FT POTENTIAL.
FT VEGETATIVE CELL WALL PROTEIN GPI.
FT 49 X 5 AA APPROXIMATE PPSFX REPEATS.
FT POLY-PRO.
FT N-LINKED (GLCNAC. . .) (POTENTIAL).
FT N-LINKED (GLCNAC. . .) (POTENTIAL).
FT N-LINKED (GLCNAC. . .) (POTENTIAL).
FT N-LINKED (GLCNAC. . .) (POTENTIAL).

```

Query Match 9.48; Score 707.5; DB 1; Length 555;

Matches 189; Conservative 66; Mismatches 243; Indels 157; Gaps 22.

393 РКЕРАРТТКЕРАРТТКЕРАРТТКСАРТТКЕРАРТТККРАРТТКЕРА 452

[illegible]

Db 40 PSPAPPS---PAPSPAPSP-----APPSPAPSPGPPSPAPSPPS 80

[illegible]

QY 453 PTTPKERTPTTPKEPARPTTKERAPTTTPKEPARPTAPKKRAPTTTPKEPARPTTPKEPARPTTK 512

[illegible][illegible]

D0 81 F-APPSPAPPSPAPPSAPSF-APPSAPPSPPSPAPPSPPAF----- 134

513 EPSPPTPKEPAPTTTKSAPTTTKEPAPTTTKSAPTTPKESPTTTKEPAPTTPKEPAPTT 572

[illegible]

135 -PCSPBP-----\$PAPPI-PPSPAPPS 154

Db 135 -PSPSP-----SPAPPLPSPAPPS 154

573 **קצת על חשיבות המחקר**

573 PKKPAPTPKEPAPTPPKPAPTPAPKEPAPTPPKLTPTTPEKL 632

100

```
Db      155 P$PPVPPSP$PPVPPSPA'P$PT-----P$S$PPVPPSPA'P$P$P$    194
```

QUESTION

QY 633 APTTPEKPAPTTPEELAPTTPEEBPTTPEEPAPTTPKAAAPNTPKKEPAPTTPKEPAPT 692

[illegible]

Db 195 APPVPPSPAPPSPSPSPAPPSPSPAPPSPVPSP-A 250

693 РКЕРАПТТПКСТАПТТЛКЕРАПТТПККАРКЕЛАПТТТКЕПСТТТСКДРАПТ 752

[illegible]

Db 251 PPSPAPPSPKPPAP-----PPPSPPPPPPRRPPFPANTMPMPPSPSPPSPA 298

QY 753 TPKGTAPTTPKEPAPTTPKGTAPTTLKEPAPTTPKKPAKELAPTTTKGPTS 812

[illegible]

Db

299 PPTPPTPPSPSPSP--VPSPAPVPPSPAPPSP---PA^SPSPPSPAP---PT^SSPSPSP 349

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

QY 813 TTSDKPAPTTPKETAPPTPKKPAPTTTPETPPPTTSEVSTPTTKEPTTIHKS 872

[illegible]

```
Db      350 SPSPSPSP-----PSPSPSPSPIPPSPPKPSPVAVKLWADDAI---- 393
```

[illegible]

873 PDESTPELSAEPT-PKALENSPKKEGVPTTKTPAATKPEMTTJAKDKTJEBDLRTJPEJTT 931

8/3 PUESIPBLSAEFI-PNALSNSPNEPGVFIKIFAAINPEMIILANDKITERDLRIIPETI 931

Db 394 ---AFDDI,NGTSTP,GSASRMVGEPTI,AGTKCK,NTI,KGMMPKPSRNP,RWGOAVESGGPTV 450

DB 394 --AFDDLNGISIRPGSASRMVGEPLDIAGIKCKGKGNLKGWMPKPSRNPWGGAVFSGGRTV 450

932 ТАДКМКЕАТТЕКМЕСКТАТТОВСТТТОДТТЕРКТТНТ КТТТТ. ДРК/ТТТТК - 989

932 TAAPKMTKETATTTETTESKITATTTTQVTSITTTQDTTFKITTLKITTLAPKVTTK-- 989

[illegible]

DB 451 GSVANVTIRVAFATEK--PALIYSIELVYNTG-----ATLRVPIANVTRSGIR 500
 QY 990 ----KTTTTEINKPEETAKKORATSKATP--PKPKPTKAPKPKPTSTKPKPT 1039
 DB 501 CPGLTYGTPTDIAGP-----TSDIATYTMNMKIAGVRINMGAKNKKPKT 545

RESULT 5
 PCLO RAT STANDARD; PRT; 5085 AA.

AC Q9JUS6; Q9JUL1;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DE 28-FEB-2003 (Rel. 41, Last annotation update)
 GN Piccolo protein (Multidomain presynaptic cytomatrix protein).
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_Taxid=10116;
 RN [1]
 RX SEQUENCE FROM N.A. (ISOFORM 2), AND INTERACTION WITH RABAC1.
 RX MEDLINE=20170257; PubMed=10707984;
 RA Fenster S.D., Chung W.J., Zhai R., Cases-Langhoff C., Voss B.,
 RA Garner A.M., Kaempf U., Kindler S., Gundelfinger E.D., Garner C.C.;
 RT "piccolo, a presynaptic zinc finger protein structurally related to
 RT bassoon.".
 RL Neuron 25:203-214 (2000).
 RN [2]
 RN SEQUENCE FROM N.A. (ISOFORM 1).
 RA Fenster S.D., Cases-Langhoff C., Gundelfinger E.D., Garner C.C.;
 RL Submitted (JAN-2000) to the EMBL/Genbank/DBJ databases.
 RN [3]
 RP CALCIUM-BINDING ACTIVITY, AND MUTAGENESIS OF ASP-4668, ASP-4674;
 RP VAL-4688; MET-4689; VAL-4690; SER-4691; GLN-4692; ASN-4693 AND
 RP ALA-4694.
 RX MEDLINE=21181819; PubMed=11285225;
 RX Geber S.H., Garcia J., Rizo J., Suedhof T.C.;
 RA "An unusual C(2)-domain in the active-zone protein piccolo:
 RT implications for Ca(2+) regulation of neurotransmitter release.";
 RL EMO J. 20:1605-1619 (2001).
 RL -1- FUNCTION: May act as a scaffolding protein involved in the
 CC organization of synaptic active zones and in synaptic vesicle
 CC trafficking (By similarity).
 CC -1- SUBUNIT: Interacts with Rabac1/Pral and profilin.
 CC -1- SUBCELLULAR LOCATION: Concentrated at presynaptic side of synaptic
 CC junctions.
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=1;
 CC IsoId=Q9JUS6-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=Q9JUS6-2; Sequence=VSP_003930; VSP_003931;
 CC -1- DOMAIN: C2 domain 1 is involved in binding calcium and
 CC phospholipids. Calcium binds with low affinity but with high
 CC specificity and induces a large conformational change.
 CC -1- SIMILARITY: Contains 2 C2 domains.
 CC -1- SIMILARITY: Contains 1 PDZ/DHR domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF138789; AAF07822.2; -;
 DR EMBL; AF227534; AAF63196.1; -;
 DR HSSP; P04107; IAZ5.
 DR GO; GO:0045202; C:synaptic junction; IDA.
 DR GO; GO:0005509; F:calcium ion binding activity; IDA.
 DR GO; GO:0005544; F:calcium-dependent phospholipid binding acti. .; IDA.

DR GO; GO:0005522; F:profilin binding activity; ISS.
 DR GO; GO:0007010; P:cytoskeleton organization and biogenesis; ISS.
 DR GO; GO:0016080; P:synaptic vesicle targeting; NAS.
 DR InterPro; IPR000008; C2.
 DR InterPro; IPR001478; PDZ.
 DR Pfam; PF00168; C2; 2.
 DR Pfam; PF00595; PDZ; 1.
 DR SMART; SM00239; C2; 2.
 DR SMART; SM00228; PDZ; 1.
 DR PROSITE; PS00499; C2_DOMAIN_1; 1.
 DR PROSITE; PS50004; C2_DOMAIN_2; 2.
 DR PROSITE; PS50106; PDZ; 1.
 KW Calcium/phospholipid-binding; Metal-binding; Zinc; Zinc-finger;
 KW Repeat; Alternative splicing.
 FT DOMAIN 372 491
 FT 12 X 10 AA TANDEM APPROXIMATE REPEATS OF
 FT P-A-K-P-Q-P-Q-P-X.
 FT C4-TYPE (POTENTIAL).
 FT C4-TYPE (POTENTIAL).
 FT POLY-PRO.
 FT PDZ.
 FT C2 DOMAIN 1.
 FT C2 DOMAIN 2.
 FT TKPTN -> SKRRK (in isoform 2).
 FT /Frid-VSP_003930.
 FT Missing (in isoform 2).
 FT /Frid-VSP_003931.
 FT D->A: COMPLETE LOSS OF CALCIUM-BINDING
 FT AND CALCIUM-DEPENDENT PHOSPHOLIPID
 FT BINDING ACTIVITY.
 FT V->S: SMALL INCREASE IN AFFINITY FOR
 FT CALCIUM.
 FT VM->SS: 10-FOLD INCREASE IN AFFINITY FOR
 FT CALCIUM.
 FT M->S: INCREASED AFFINITY FOR CALCIUM.
 FT VV->SS: 10-FOLD INCREASE IN AFFINITY FOR
 FT CALCIUM.
 FT ON->AA: MODERATE INCREASE IN AFFINITY FOR
 FT CALCIUM.
 FT A->S: NO EFFECT ON CALCIUM-BINDING
 FT ACTIVITY.
 SQ SEQUENCE 5085 AA; 552702 MW; 5A1BB543201A7450 CRC64;
 Query Match 9.3%; Score 703.5; DB 1; Length 5085;
 Best Local Similarity 24.1%; Pred. No. 1.3e-19;
 Matches 313; Conservative 118; Mismatches 437; Indels 431; Gaps 60;

QY 113 PPSKKAPPEGASOTIKSTTKRSP-KRP-----NKKTKVYSEB----- 153
 DB 68 PPAABESPMNRKKELDSSQAPQPGKPPDGRPTGRLSKRTTDFRSQKLPGRSPS 127
 QY 154 ---ITTEHVSSENGESSSSSS-----SSSSSTIWKIS-----SKNSAANRELQK 196
 DB 128 TISLSEKSRIDFEBYKSSMMPGFSDVNLVAVSSVVKNFDFILSDSEAOEFTTK 187
 QY 197 KLVKDKNKKRNTKKKPPKPPVVDKAGSLDNGPKVTTPTSTTQHNKYSTSKITTK 256
 DB 188 KQKVV--QKEGKSGNAKPL-----Q 208
 QY 257 PINRPSLPPNSDTSKETSILVNKETTVETKETTNNKQTSIDGKETSAAKQSIKKT 316
 DB 209 QPSKPP-IPKQGGYKVIQDSSPKVSSQAKVAPQABGTGK-----PSQOSPQOT 261
 QY 317 SAKDLAFTSKVLAAP-TPKATTTKGPALTPKPEPTTPKPEASTTPKEPTPTTIKSAP 375
 DB 262 PAQASGPKVPAQPGSAKATVQGPDA-KSPAIPA-GTGKSPAQPAKTRGQQAQLEKT 319
 QY 376 TTKEPAPTTKKAPTP-----KEPA-PTTKEPAPTTPKEPA-PTTKEPA 422
 DB 320 SSSQPGP-----KSIAQPGHGKPLGPKVKSAPQPGAKHPA-----QDPGPTAKVPGP 372

QY	423	TTTGAAPTTKPKBAPPTTKKAPAPLT---	TPKBPAPTTKBPPTTKPKBAPPTTKBPAPTTTK	480	
DB	373	TKT---	PAQSGEPKTPAQCPGPTKSPQCPRIAPKQPOCPATK--	TOPOGAPAPQPO 428	
QY	481	BPAPTPAKKBPAPTTKPK-BEAPPTPKBPAPTTTKBPSPPTTKBPBPAPTTTK---	SAPPTTK	536	
DB	429	QPAKAPPOQCPPTPAKPOQPPRTPAKPOP---	QPPATIPPOQPPRTATIPPHQPOELAP	485	
QY	537	PAPPTTKGAPTT---	EKBPSTTTKE-PAPPTTKBPAP---	TT-----PKX----- 575	
DB	486	SAQCPPTKSGIVTGRPLQPPRTSAQAOTPAQGLSKITICPLCTTELLLI	PEKANFNCTG	545	
QY	576	-----	PAPTPKE-----	PAPTPKBPAPTTTKBPAP 602	
DB	546	ECQSTVCSLCGFNPNNHLEIKEMWCLNCOMQORALGDLAALIPSSP-	QPTPAATAPTA	604	
QY	603	TAPKEBAPT---	TPKETAPTTEKCLPTTPPEKLAPTTPKEBPAPTTKEBAPTTKEBAPTTP	658	
DB	605	TASKSPVPSQOASPKPEPSS--	KQDSPKALSKKPPPEPKPP-----	PEPKPPPEPKPPDPL 658	
QY	659	-----	TTPEEP-APPTPKAADNTPKBPAPTTKBP-APPTPKBPAPT---	699	
DB	659	VKQPLHGPRTAPATAPQLPVAAALPEBAPPKPEPSGRLPEQAKAPVGDVBPQPMETRAD		718	
QY	700	-----	TPKETAPTTPKGAAPTTLK-----	718	
DB	719	IQSSSTTKRDLISSQVQOAVKTASPLKTDASKPQSPPEPGEKTPPLDSDKMPAPAD		778	
QY	719	-----	BPAPTTPEKBPAPKELAPTTTKEPRTSDKAPAPTTPKGATAPTTKBPAPTTPKBP	773	
DB	779	SKIIQCPGSGSSKD-PKHIDPIQKK-----	DEKKAKQPKQ-----	SKPEPKAPVPK-- 824	
QY	774	APTTKEGTAAPTTKEBPAPTTPKKBPAPKELAPTTKQPTS--	TSDKBPAPTTPKETAPTTK	832	
DB	825	GSPTSGTRPTA-GQAAPPSQCPKQEQSRRSLNLGGITAPKSGPPTTPQET--	VTGK	881	
QY	833	-----	BPAPTTPKBPAPTTPEEP---	PTTSVSTPTTT 863	
DB	882	LEFGGASIFSQASNLSTAGQOGPHQOTGPAPASKQAPTPPSQSPAPAQGPASKTGOLPAP		941	
QY	864	KEPTTIHNSPD-----	ESTPELSAEPPTKALENSKEPVGPTTTP-----	AATK 908	
DB	942	AKATIVKKEAKAAALENLESKEPEQA--	PTAKTTEKQKKPPPAKVGPSPSEPEKAPAPAK	999	
QY	909	PEMTTAKDK---	TTERDL-----	RTTPETT----- 931	
DB	1000	PDKTTKRPACPLCTELNLGSGQBPNNFTCTCBCKNQVNLGFGNTPHLTELQEWLCIN		105	
QY	932	-----	TAAPKTKETATTEKTESKITATTTOVT-----	STTQ 966	
DB	1060	COTORAISQGLDMGKMPAPSPGPKAPSPAPAPAPSSQCKTPIGTQVKGKKKEAEGTKEA		111	
QY	967	DTTPKTIITLKTTLAPKYTTTKITITTEIMNKKEPETAAPKORAINSKATTPKQPKAPK		102	
DB	1120	KPVPEKETASISIKT--	PMVYTTQCKL-----	ESSGCKK-----	SKYSALPEKXSE 116
QY	1027	APKKESTTKPKPTMPVRKP-----	KTTPTPKMTSTWMEINLPTSRILAIAMQLTTTPRQ	108	
DB	1164	BEKALISADKKEKKPPAPAEKRPJLEBKCPJIVDKLP--	PEAKPLS-----	120	
QY	1082	TPNSKLVEVNPKESDAGAEGETPHMLRPVAVPMPEVTP		1120	
DB	1206	-----	SEGGEKHEILLAKAHQIPEEP	1226	

RESULT 6

PCLO_MOUSE

STANDARD;

PRT; 5038 AA.

Q9QYX7, Q9QYX6, Q9QZU0;

28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Plcoco protein (preynapptic cytomatrix protein) (Aczonin) (Brain-

DE derived H1MN protein).
GN PC10 OR ACZ.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
[1]
RP SEQUENCE FROM N.A., SUBCELLULAR LOCATION, ALTERNATIVE SPLICING,
RT TISSUE SPECIFICITY, AND INTERACTION WITH PROFILIN.
RC TISSUE=Brain;
RX MEDLINE=99439764; Pubmed=10508862;
RA Wang X., Kibschull M., Laue M.M., Lichte B., Petrasch-Parwez E.,
RT "Accomn, a 550-kd putative scaffolding protein of presynaptic active
RT zones, shares homology regions with rim and bassoon and binds
RT profilin.";
RL J. Cell Biol. 147:151-162(1999).
RN [2]
RP REVISIONS.
RC TISSUE=Brain;
RA Kilimann M.;
RN Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
[3]
RP SEQUENCE OF 4502-4682 FROM N.A.
RC TISSUE=Brain;
RA Huang W., Jin W., Huang C., Chen B., Zhang J., Ju G.;
RN Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: May act as a scaffolding protein involved in the
CC organization of synaptic active zones and in synaptic vesicle
CC trafficking.
CC -1- SUBUNIT: Interacts with Rabac1/Pral and profilin.
CC -1- SUBCELLULAR LOCATION: Concentrated at the presynaptic side of
CC synaptic junctions.
CC -1- ALTERNATIVE PRODUCTS:
CC Event-Alternative splicing; Named isoforms=2;
CC Name=1;
CC IsoId=Q9QYX7-1; Sequence=displayed;
CC Name=2;
CC IsoId=Q9QYX7-2; Sequence=VSP_003928, VSP_003929;
CC -1- TISSUE SPECIFICITY: Highly expressed in brain. Low levels found in
CC stomach. Not detected in other tissues analyzed including adrenal
CC gland, testis and pancreas.
CC -1- DOMAIN: C2 domain 1 is involved in binding calcium and
CC phospholipids. Calcium binds with low affinity but with high
CC specificity and induces a large conformational change.
CC -1- SIMILARITY: Contains 2 C2 domains.
CC -1- SIMILARITY: Contains 1 PDZ/DHR domain.

CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC
DR EMBL; Y19185; CAB60731.2; -;
DR EMBL; Y19186; CAB60732.2; -;
DR EMBL; AF181269; AAD55786.2; -;
DR HSSP; P04410; 1A25.
DR MGD; MGI:1349390; PclO.
DR GO; GO:0045202; C:synaptic junction; IDA.
DR GO; GO:0005509; F:calcium ion binding activity; ISS.
DR GO; GO:0005549; F:calcium-dependent phospholipid binding acti. . .; ISS.
DR GO; GO:0005552; F:profilin binding activity; IDA.
DR GO; GO:0007010; F:cytoskeleton organization and biogenesis; IDA.
DR GO; GO:0016080; P:synaptic vesicle targeting; NAS.
DR InterPro; IPR000008; C2.
DR Pfam; PF00168; C2; 2.
DR Pfam; PF00595; PDZ; 1.
DR SMART; SM00239; C2; 2.
DR SMART; SM00238; PDZ; 1.

RC TISSUE=Placenta;
RX MEDLINE=9716151; PubMed=9013550;
RA Desseyn J.-L., Guyonnet-Duperrat V., Porchet N., Aubert J.-P.,
RA Laine A.,
RT "Human mucin gene MUC5B, the 10.7 kb large central exon encodes
RT various alternate subdomains resulting in a super-repeat. Structural
RT evidence for a 11p15.5 gene family.";
RL J. Biol. Chem. 272:3168-3176(1997).
RN [5]
RP SEQUENCE OF 4057-4480 FROM N.A.
RC TISSUE=Salivary gland;
RX MEDLINE=97292540; PubMed=9147051;
RA Nielsen P.A., Bennett E.P., Wandall H.H., Therkildsen M.H.,
RA Handhal J., Clausen H.,
RT "Identification of a major human high molecular weight salivary mucin
RT (MG1) as tracheobronchial mucin MUC5B.";
RL Glycobiology 7:413-419(1997).
RN [6]
RP SEQUENCE OF 4721-5703 FROM N.A.
RC TISSUE=Gall bladder;
RX MEDLINE=97293229; PubMed=9164870;
RA Keates A.C., Nunes D.P., Aldhal N.H., Troxler R.F., Offner G.D.,
RT "Molecular cloning of a major human gall bladder mucin: complete C-
RT terminal sequence and genomic organization of MUC5B.";
RL Biochem. J. 324:295-303(1997).
RN [7]
RP SEQUENCE OF 4809-5687 FROM N.A.
RC TISSUE=Sublingual gland;
RX MEDLINE=96125355; PubMed=8554565;
RA Troxler R.F., Offner G.D., Zhang F., Iontcheva I., Oppenheim F.G.,
RT "Molecular cloning of a novel high molecular weight mucin (MG1)
RT from human sublingual gland.";
RL Biochem. Biophys. Res. Commun. 217:1112-1119(1995).
RN [8]
RP SEQUENCE OF 4859-5703 FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=97347489; PubMed=9201995;
RA Desseyn J.-L., Aubert J.-P., Porchet N., Laine A.,
RT "Genomic organization of the 3 region of the human MUC5B mucin";
RL J. Biol. Chem. 272:16873-16883(1997).
CC -1- FUNCTION: Salivary mucin that is thought to contribute to the
CC lubricating and viscoelastic properties of whole saliva.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: Expressed mainly in bronchus glands and also
CC in submaxillary glands, endocervix, gall bladder, and pancreas.
CC -1- PTM: Highly glycosylated.
CC -1- SIMILARITY: Contains 1 TIL (Trypsin inhibitory-like) domain.
CC -1- SIMILARITY: Contains 3 WFCD domains.
CC -1- SIMILARITY: Contains 4 WFCD domains.
CC -1- SIMILARITY: Contains 1 C-terminal cysteine knot-like (CTCK) domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, AF107890; AAC33673.1; -
DR EMBL, AF086604; AAC67545.1; -
DR EMBL, AF004862; CAA0657.1; -
DR EMBL, Z72496; CAA96577.1; -
DR EMBL, X74955; CAA52910.1; -
DR EMBL, U63836; AAB61398.1; -
DR EMBL, U78554; AAC51344.1; -
DR EMBL, U78552; AAC51344.1; JOINED.
DR EMBL, U78553; AAC51344.1; JOINED.
DR EMBL, U78551; AAC51343.1; -
DR EMBL, U95031; AAB65151.1; -
DR EMBL, Y09788; CAA70926.1; -
DR GeneW, HGNC:7516; MUC5B.
DR MIM, 600770; -.

DR GO; GO:0005515; F:protein binding activity; IPI.
DR InterPro: IPR006208; Cys_knot.
DR InterPro: IPR006207; Cys_knot.
DR InterPro: IPR002919; TIL_CysRich.
DR InterPro: IPR006552; VC_out.
DR InterPro: IPR01007; WF_C.
DR InterPro: IPR01846; WF_D.
DR Pfam: PF00007; Cys_knot_1.
DR Pfam: PF01826; TIL_1.
DR Pfam: PF00093; WFC_1.
DR Pfam: PF00094; WFC_4.
DR SMART: SM00214; WFC_6.
DR SMART: SM00215; WFC_out; 4.
DR SMART: SM00216; WMD; 4.
DR PROSITE: PS01185; CTCK_1.1.
DR PROSITE: PS01225; CTCK_2.1.
DR PROSITE: PS01208; WFC_1; 2.
DR PROSITE: PS0184; WFC_2; 2.
DR GlycoProtein; Repeat; Signal; Polymorphism.
FT CHAIN 1 25
FT SIGNAL 1 25
FT CHAIN 26 5703
FT SIGNAL 26 5703
FT CHAIN 27 77
FT SIGNAL 27 77
FT CHAIN 28 329
FT SIGNAL 28 329
FT CHAIN 29 426
FT SIGNAL 29 426
FT CHAIN 30 858
FT SIGNAL 30 858
FT CHAIN 31 896
FT SIGNAL 31 896
FT CHAIN 32 1457
FT SIGNAL 32 1457
FT CHAIN 33 1609
FT SIGNAL 33 1609
FT CHAIN 34 5005
FT SIGNAL 34 5005
FT CHAIN 35 5353
FT SIGNAL 35 5353
FT CHAIN 36 5462
FT SIGNAL 36 5462
FT CHAIN 37 5594
FT SIGNAL 37 5594
FT CHAIN 38 5645
FT SIGNAL 38 5645
FT CHAIN 39 145
FT SIGNAL 39 145
FT CHAIN 40 201
FT SIGNAL 40 201
FT CHAIN 41 254
FT SIGNAL 41 254
FT CHAIN 42 402
FT SIGNAL 42 402
FT CHAIN 43 516
FT SIGNAL 43 516
FT CHAIN 44 806
FT SIGNAL 44 806
FT CHAIN 45 930
FT SIGNAL 45 930
FT CHAIN 46 1277
FT SIGNAL 46 1277
FT CHAIN 47 1293
FT SIGNAL 47 1293
FT CHAIN 48 1557
FT SIGNAL 48 1557
FT CHAIN 49 1775
FT SIGNAL 49 1775
FT CHAIN 50 2192
FT SIGNAL 50 2192
FT CHAIN 51 2721
FT SIGNAL 51 2721
FT CHAIN 52 3419
FT SIGNAL 52 3419
FT CHAIN 53 3948
FT SIGNAL 53 3948
FT CHAIN 54 4745
FT SIGNAL 54 4745
FT CHAIN 55 4901
FT SIGNAL 55 4901
FT CHAIN 56 4958
FT SIGNAL 56 4958
FT CHAIN 57 4965
FT SIGNAL 57 4965
FT CHAIN 58 4987
FT SIGNAL 58 4987
FT CHAIN 59 5037
FT SIGNAL 59 5037
FT CHAIN 60 5052
FT SIGNAL 60 5052
FT CHAIN 61 5156
FT SIGNAL 61 5156
FT CHAIN 62 5427
FT SIGNAL 62 5427
FT CHAIN 63 5467
FT SIGNAL 63 5467
FT CHAIN 64 5506
FT SIGNAL 64 5506
FT CHAIN 65 5507
FT SIGNAL 65 5507
FT CHAIN 66 5543
FT SIGNAL 66 5543
FT CHAIN 67 5553
FT SIGNAL 67 5553
FT CHAIN 68 5604
FT SIGNAL 68 5604
FT CHAIN 69 5618
FT SIGNAL 69 5618
FT CHAIN 70 5662
FT SIGNAL 70 5662
FT CHAIN 71 5137
FT SIGNAL 71 5137
FT CHAIN 72 34
FT SIGNAL 72 34
FT CHAIN 73 95
FT SIGNAL 73 95
FT CHAIN 74 100
FT SIGNAL 74 100
FT CHAIN 75 34
FT SIGNAL 75 34
FT CHAIN 76 95
FT SIGNAL 76 95
FT CHAIN 77 100
FT SIGNAL 77 100
FT CHAIN 78 34
FT SIGNAL 78 34
FT CHAIN 79 95
FT SIGNAL 79 95
FT CHAIN 80 100
FT SIGNAL 80 100
FT CHAIN 81 34
FT SIGNAL 81 34
FT CHAIN 82 95
FT SIGNAL 82 95
FT CHAIN 83 100
FT SIGNAL 83 100
FT CHAIN 84 34
FT SIGNAL 84 34
FT CHAIN 85 95
FT SIGNAL 85 95
FT CHAIN 86 100
FT SIGNAL 86 100
FT CHAIN 87 34
FT SIGNAL 87 34
FT CHAIN 88 95
FT SIGNAL 88 95
FT CHAIN 89 100
FT SIGNAL 89 100
FT CHAIN 90 34
FT SIGNAL 90 34
FT CHAIN 91 95
FT SIGNAL 91 95
FT CHAIN 92 100
FT SIGNAL 92 100
FT CHAIN 93 34
FT SIGNAL 93 34
FT CHAIN 94 95
FT SIGNAL 94 95
FT CHAIN 95 100
FT SIGNAL 95 100
FT CHAIN 96 34
FT SIGNAL 96 34
FT CHAIN 97 95
FT SIGNAL 97 95
FT CHAIN 98 100
FT SIGNAL 98 100
FT CHAIN 99 34
FT SIGNAL 99 34
FT CHAIN 100 95
FT SIGNAL 100 95
FT CHAIN 101 100
FT SIGNAL 101 100
FT CHAIN 102 34
FT SIGNAL 102 34
FT CHAIN 103 95
FT SIGNAL 103 95
FT CHAIN 104 100
FT SIGNAL 104 100
FT CHAIN 105 34
FT SIGNAL 105 34
FT CHAIN 106 95
FT SIGNAL 106 95
FT CHAIN 107 100
FT SIGNAL 107 100
FT CHAIN 108 34
FT SIGNAL 108 34
FT CHAIN 109 95
FT SIGNAL 109 95
FT CHAIN 110 100
FT SIGNAL 110 100
FT CHAIN 111 34
FT SIGNAL 111 34
FT CHAIN 112 95
FT SIGNAL 112 95
FT CHAIN 113 100
FT SIGNAL 113 100
FT CHAIN 114 34
FT SIGNAL 114 34
FT CHAIN 115 95
FT SIGNAL 115 95
FT CHAIN 116 100
FT SIGNAL 116 100
FT CHAIN 117 34
FT SIGNAL 117 34
FT CHAIN 118 95
FT SIGNAL 118 95
FT CHAIN 119 100
FT SIGNAL 119 100
FT CHAIN 120 34
FT SIGNAL 120 34
FT CHAIN 121 95
FT SIGNAL 121 95
FT CHAIN 122 100
FT SIGNAL 122 100
FT CHAIN 123 34
FT SIGNAL 123 34
FT CHAIN 124 95
FT SIGNAL 124 95
FT CHAIN 125 100
FT SIGNAL 125 100
FT CHAIN 126 34
FT SIGNAL 126 34
FT CHAIN 127 95
FT SIGNAL 127 95
FT CHAIN 128 100
FT SIGNAL 128 100
FT CHAIN 129 34
FT SIGNAL 129 34
FT CHAIN 130 95
FT SIGNAL 130 95
FT CHAIN 131 100
FT SIGNAL 131 100
FT CHAIN 132 34
FT SIGNAL 132 34
FT CHAIN 133 95
FT SIGNAL 133 95
FT CHAIN 134 100
FT SIGNAL 134 100
FT CHAIN 135 34
FT SIGNAL 135 34
FT CHAIN 136 95
FT SIGNAL 136 95
FT CHAIN 137 100
FT SIGNAL 137 100
FT CHAIN 138 34
FT SIGNAL 138 34
FT CHAIN 139 95
FT SIGNAL 139 95
FT CHAIN 140 100
FT SIGNAL 140 100
FT CHAIN 141 34
FT SIGNAL 141 34
FT CHAIN 142 95
FT SIGNAL 142 95
FT CHAIN 143 100
FT SIGNAL 143 100
FT CHAIN 144 34
FT SIGNAL 144 34
FT CHAIN 145 95
FT SIGNAL 145 95
FT CHAIN 146 100
FT SIGNAL 146 100
FT CHAIN 147 34
FT SIGNAL 147 34
FT CHAIN 148 95
FT SIGNAL 148 95
FT CHAIN 149 100
FT SIGNAL 149 100
FT CHAIN 150 34
FT SIGNAL 150 34
FT CHAIN 151 95
FT SIGNAL 151 95
FT CHAIN 152 100
FT SIGNAL 152 100
FT CHAIN 153 34
FT SIGNAL 153 34
FT CHAIN 154 95
FT SIGNAL 154 95
FT CHAIN 155 100
FT SIGNAL 155 100
FT CHAIN 156 34
FT SIGNAL 156 34
FT CHAIN 157 95
FT SIGNAL 157 95
FT CHAIN 158 100
FT SIGNAL 158 100
FT CHAIN 159 34
FT SIGNAL 159 34
FT CHAIN 160 95
FT SIGNAL 160 95
FT CHAIN 161 100
FT SIGNAL 161 100
FT CHAIN 162 34
FT SIGNAL 162 34
FT CHAIN 163 95
FT SIGNAL 163 95
FT CHAIN 164 100
FT SIGNAL 164 100
FT CHAIN 165 34
FT SIGNAL 165 34
FT CHAIN 166 95
FT SIGNAL 166 95
FT CHAIN 167 100
FT SIGNAL 167 100
FT CHAIN 168 34
FT SIGNAL 168 34
FT CHAIN 169 95
FT SIGNAL 169 95
FT CHAIN 170 100
FT SIGNAL 170 100
FT CHAIN 171 34
FT SIGNAL 171 34
FT CHAIN 172 95
FT SIGNAL 172 95
FT CHAIN 173 100
FT SIGNAL 173 100
FT CHAIN 174 34
FT SIGNAL 174 34
FT CHAIN 175 95
FT SIGNAL 175 95
FT CHAIN 176 100
FT SIGNAL 176 100
FT CHAIN 177 34
FT SIGNAL 177 34
FT CHAIN 178 95
FT SIGNAL 178 95
FT CHAIN 179 100
FT SIGNAL 179 100
FT CHAIN 180 34
FT SIGNAL 180 34
FT CHAIN 181 95
FT SIGNAL 181 95
FT CHAIN 182 100
FT SIGNAL 182 100
FT CHAIN 183 34
FT SIGNAL 183 34
FT CHAIN 184 95
FT SIGNAL 184 95
FT CHAIN 185 100
FT SIGNAL 185 100
FT CHAIN 186 34
FT SIGNAL 186 34
FT CHAIN 187 95
FT SIGNAL 187 95
FT CHAIN 188 100
FT SIGNAL 188 100
FT CHAIN 189 34
FT SIGNAL 189 34
FT CHAIN 190 95
FT SIGNAL 190 95
FT CHAIN 191 100
FT SIGNAL 191 100
FT CHAIN 192 34
FT SIGNAL 192 34
FT CHAIN 193 95
FT SIGNAL 193 95
FT CHAIN 194 100
FT SIGNAL 194 100
FT CHAIN 195 34
FT SIGNAL 195 34
FT CHAIN 196 95
FT SIGNAL 196 95
FT CHAIN 197 100
FT SIGNAL 197 100
FT CHAIN 198 34
FT SIGNAL 198 34
FT CHAIN 199 95
FT SIGNAL 199 95
FT CHAIN 200 100
FT SIGNAL 200 100
FT CHAIN 201 34
FT SIGNAL 201 34
FT CHAIN 202 95
FT SIGNAL 202 95
FT CHAIN 203 100
FT SIGNAL 203 100
FT CHAIN 204 34
FT SIGNAL 204 34
FT CHAIN 205 95
FT SIGNAL 205 95
FT CHAIN 206 100
FT SIGNAL 206 100
FT CHAIN 207 34
FT SIGNAL 207 34
FT CHAIN 208 95
FT SIGNAL 208 95
FT CHAIN 209 100
FT SIGNAL 209 100
FT CHAIN 210 34
FT SIGNAL 210 34
FT CHAIN 211 95
FT SIGNAL 211 95
FT CHAIN 212 100
FT SIGNAL 212 100
FT CHAIN 213 34
FT SIGNAL 213 34
FT CHAIN 214 95
FT SIGNAL 214 95
FT CHAIN 215 100
FT SIGNAL 215 100
FT CHAIN 216 34
FT SIGNAL 216 34
FT CHAIN 217 95
FT SIGNAL 217 95
FT CHAIN 218 100
FT SIGNAL 218 100
FT CHAIN 219 34
FT SIGNAL 219 34
FT CHAIN 220 95
FT SIGNAL 220 95
FT CHAIN 221 100
FT SIGNAL 221 100
FT CHAIN 222 34
FT SIGNAL 222 34
FT CHAIN 223 95
FT SIGNAL 223 95
FT CHAIN 224 100
FT SIGNAL 224 100
FT CHAIN 225 34
FT SIGNAL 225 34
FT CHAIN 226 95
FT SIGNAL 226 95
FT CHAIN 227 100
FT SIGNAL 227 100
FT CHAIN 228 34
FT SIGNAL 228 34
FT CHAIN 229 95
FT SIGNAL 229 95
FT CHAIN 230 100
FT SIGNAL 230 100
FT CHAIN 231 34
FT SIGNAL 231 34
FT CHAIN 232 95
FT SIGNAL 232 95
FT CHAIN 233 100
FT SIGNAL 233 100
FT CHAIN 234 34
FT SIGNAL 234 34
FT CHAIN 235 95
FT SIGNAL 235 95
FT CHAIN 236 100
FT SIGNAL 236 100
FT CHAIN 237 34
FT SIGNAL 237 34
FT CHAIN 238 95
FT SIGNAL 238 95
FT CHAIN 239 100
FT SIGNAL 239 100
FT CHAIN 240 34
FT SIGNAL 240 34
FT CHAIN 241 95
FT SIGNAL 241 95
FT CHAIN 242 100
FT SIGNAL 242 100
FT CHAIN 243 34
FT SIGNAL 243 34
FT CHAIN 244 95
FT SIGNAL 244 95
FT CHAIN 245 100
FT SIGNAL 245 100
FT CHAIN 246 34
FT SIGNAL 246 34
FT CHAIN 247 95
FT SIGNAL 247 95
FT CHAIN 248 100
FT SIGNAL 248 100
FT CHAIN 249 34
FT SIGNAL 249 34
FT CHAIN 250 95
FT SIGNAL 250 95
FT CHAIN 251 100
FT SIGNAL 251 100
FT CHAIN 252 34
FT SIGNAL 252 34
FT CHAIN 253 95
FT SIGNAL 253 95
FT CHAIN 254 100
FT SIGNAL 254 100
FT CHAIN 255 34
FT SIGNAL 255 34
FT CHAIN 256 95
FT SIGNAL 256 95
FT CHAIN 257 100
FT SIGNAL 257 100
FT CHAIN 258 34
FT SIGNAL 258 34
FT CHAIN 259 95
FT SIGNAL 259 95
FT CHAIN 260 100
FT SIGNAL 260 100
FT CHAIN 261 34
FT SIGNAL 261 34
FT CHAIN 262 95
FT SIGNAL 262 95
FT CHAIN 263 100
FT SIGNAL 263 100
FT CHAIN 264 34
FT SIGNAL 264 34
FT CHAIN 265 95
FT SIGNAL 265 95
FT CHAIN 266 100
FT SIGNAL 266 100
FT CHAIN 267 34
FT SIGNAL 267 34
FT CHAIN 268 95
FT SIGNAL 268 95
FT CHAIN 269 100
FT SIGNAL 269 100
FT CHAIN 270 34
FT SIGNAL 270 34
FT CHAIN 271 95
FT SIGNAL 271 95
FT CHAIN 272 100
FT SIGNAL 272 100
FT CHAIN 273 34
FT SIGNAL 273 34
FT CHAIN 274 95
FT SIGNAL 274 95
FT CHAIN 275 100
FT SIGNAL 275 100
FT CHAIN 276 34
FT SIGNAL 276 34
FT CHAIN 277 95
FT SIGNAL 277 95
FT CHAIN 278 100
FT SIGNAL 278 100
FT CHAIN 279 34
FT SIGNAL 279 34
FT CHAIN 280 95
FT SIGNAL 280 95
FT CHAIN 281 100
FT SIGNAL 281 100
FT CHAIN 282 34
FT SIGNAL 282 34
FT CHAIN 283 95
FT SIGNAL 283 95
FT CHAIN 284 100
FT SIGNAL 284 100
FT CHAIN 285 34
FT SIGNAL 285 34
FT CHAIN 286 95
FT SIGNAL 286 95
FT CHAIN 287 100
FT SIGNAL 287 100
FT CHAIN 288 34
FT SIGNAL 288 34
FT CHAIN 289 95
FT SIGNAL 289 95
FT CHAIN 290 100
FT SIGNAL 290 100
FT CHAIN 291 34
FT SIGNAL 291 34
FT CHAIN 292 95
FT SIGNAL 292 95
FT CHAIN 293 100
FT SIGNAL 293 100
FT CHAIN 294 34
FT SIGNAL 294 34
FT CHAIN 295 95
FT SIGNAL 295 95
FT CHAIN 296 100
FT SIGNAL 296 100
FT CHAIN 297 34
FT SIGNAL 297 34
FT CHAIN 298 95
FT SIGNAL 298 95
FT CHAIN 299 100
FT SIGNAL 299 100
FT CHAIN 300 34
FT SIGNAL 300 34
FT CHAIN 301 95
FT SIGNAL 301 95
FT CHAIN 302 100
FT SIGNAL 302 100
FT CHAIN 303 34
FT SIGNAL 303 34
FT CHAIN 304 95
FT SIGNAL 304 95
FT CHAIN 305 100
FT SIGNAL 305 100
FT CHAIN 306 34
FT SIGNAL 306 34
FT CHAIN 307 95
FT SIGNAL 307 95
FT CHAIN 308 100
FT SIGNAL 308 100
FT CHAIN 309 34
FT SIGNAL 309 34
FT CHAIN 310 95
FT SIGNAL 310 95
FT CHAIN 311 100
FT SIGNAL 311 100
FT CHAIN 312 34
FT SIGNAL 312 34
FT CHAIN 313 95
FT SIGNAL 313 95
FT CHAIN 314 100
FT SIGNAL 314 100
FT CHAIN 315 34
FT SIGNAL 315 34
FT CHAIN 316 95
FT SIGNAL 316 95
FT CHAIN 317 100
FT SIGNAL 317 100
FT CHAIN 318 34
FT SIGNAL 318 34
FT CHAIN 319 95
FT SIGNAL 319 95
FT CHAIN 320 100
FT SIGNAL 320 100
FT CHAIN 321 34
FT SIGNAL 321 34
FT CHAIN 322 95
FT SIGNAL 322 95
FT CHAIN 323 100
FT SIGNAL 323 100
FT CHAIN 324 34
FT SIGNAL 324 34
FT CHAIN 325 95
FT SIGNAL 325 95
FT CHAIN 326 100
FT SIGNAL 326 100
FT CHAIN 327 34
FT SIGNAL 327 34
FT CHAIN 328 95
FT SIGNAL 328 95
FT CHAIN 329 100
FT SIGNAL 329 100
FT CHAIN 330 34
FT SIGNAL 330 34
FT CHAIN 331 95
FT SIGNAL 331 95
FT CHAIN 332 100
FT SIGNAL 332 100
FT CHAIN 333 34
FT SIGNAL 333 34
FT CHAIN 334 95
FT SIGNAL 334 95
FT CHAIN 335 100
FT SIGNAL 335 100
FT CHAIN 336 34
FT SIGNAL 336 34
FT CHAIN 337 95
FT SIGNAL 337 95
FT CHAIN 338 100
FT SIGNAL 338 100
FT CHAIN 339 34
FT SIGNAL 339 34
FT CHAIN 340 95
FT SIGNAL 340 95
FT CHAIN 341 100
FT SIGNAL 341 100
FT CHAIN 342 34
FT SIGNAL 342 34
FT CHAIN 343 95
FT SIGNAL 343 95
FT CHAIN 344 100
FT SIGNAL 344 100
FT CHAIN 345 34
FT SIGNAL 345 34
FT CHAIN 346 95
FT SIGNAL 346 95
FT CHAIN 347 100
FT SIGNAL 347 100
FT CHAIN 348 34
FT SIGNAL 348 34
FT CHAIN 349 95
FT SIGNAL 349 95
FT CHAIN 350 100
FT SIGNAL 350 100
FT CHAIN 351 34
FT SIGNAL 351 34
FT CHAIN 352 95
FT SIGNAL 352 95
FT CHAIN 353 100
FT SIGNAL 353 100
FT CHAIN 354 34
FT SIGNAL 354 34
FT CHAIN 355 95
FT SIGNAL 355 95
FT CHAIN 356 100
FT SIGNAL 356 100
FT CHAIN 357 34
FT SIGNAL 357 34
FT CHAIN 358 95
FT SIGNAL 358 95
FT CHAIN 359 100
FT SIGNAL 359 100
FT CHAIN 360 34
FT SIGNAL 360 34
FT CHAIN 361 95
FT SIGNAL 361 95
FT CHAIN 362 100
FT SIGNAL 362 100
FT CHAIN 363 34
FT SIGNAL 363 34
FT CHAIN 364 95
FT SIGNAL 364 95
FT CHAIN 365 100
FT SIGNAL 365 100
FT CHAIN 366 34
FT SIGNAL 366 34
FT CHAIN 367 95
FT SIGNAL 367 95
FT CHAIN 368 100
FT SIGNAL 368 100
FT CHAIN 369 34
FT SIGNAL 369 34
FT CHAIN 370 95
FT SIGNAL 370 95
FT CHAIN 371 100
FT SIGNAL 371 100
FT CHAIN 372 34
FT SIGNAL 372 34
FT CHAIN 373 95
FT SIGNAL 373 95
FT CHAIN 374 100
FT SIGNAL 374 100
FT CHAIN 375 34
FT SIGNAL 375 34
FT CHAIN 376 95
FT SIGNAL 376 95
FT CHAIN 377 100
FT SIGNAL 377 100
FT CHAIN 378 34
FT SIGNAL 378 34
FT CHAIN 379 95
FT SIGNAL 379 95
FT CHAIN 380 100
FT SIGNAL 380 100
FT CHAIN 381 34
FT SIGNAL 381 34
FT CHAIN 382 95
FT SIGNAL 382 95
FT CHAIN 383 100
FT SIGNAL 383 100
FT CHAIN 384 34
FT SIGNAL 384 34
FT CHAIN 385 95
FT SIGNAL 385 95
FT CHAIN 386 100
FT SIGNAL 386 100
FT CHAIN 387 34
FT SIGNAL 387 34
FT CHAIN 388 95
FT SIGNAL 388 95
FT CHAIN 389 100
FT SIGNAL 389 100
FT CHAIN 390 34
FT SIGNAL 390 34
FT CHAIN 391 95
FT SIGNAL 391 95
FT CHAIN 392 100
FT SIGNAL 392 100
FT CHAIN 393 34
FT SIGNAL 393 34
FT CHAIN 394 95
FT SIGNAL 394 95
FT CHAIN 395 100
FT SIGNAL 395 100
FT CHAIN 396 34
FT SIGNAL 396 34
FT CHAIN 397 95
FT SIGNAL 397 95
FT CHAIN 398 100
FT SIGNAL 398 100
FT CHAIN 399 34
FT SIGNAL 399 34
FT CHAIN 400 95
FT SIGNAL 400 95
FT CHAIN 401 100
FT SIGNAL 401 100
FT CHAIN 402 34
FT SIGNAL 402 34
FT CHAIN 403 95
FT SIGNAL 403 95
FT CHAIN 404 100
FT SIGNAL 404 100
FT CHAIN 405 34
FT SIGNAL 405 34
FT CHAIN 406 95
FT SIGNAL 406 95
FT CHAIN 407 100
FT SIGNAL 407 100
FT CHAIN 408 34
FT SIGNAL 408 34
FT CHAIN 409 95
FT SIGNAL 409 95
FT CHAIN 410 100
FT SIGNAL 410 100
FT CHAIN 411 34
FT SIGNAL 411 34
FT CHAIN 412 95
FT SIGNAL 412 95
FT CHAIN 413 100
FT SIGNAL 413 100
FT CHAIN 414 34
FT SIGNAL 414 34
FT CHAIN 415 95
FT SIGNAL 415 95
FT CHAIN 416 100
FT SIGNAL 416 100
FT CHAIN 417 34
FT SIGNAL 417 34
FT CHAIN 418 95
FT SIGNAL 418 95
FT CHAIN 419 100
FT SIGNAL 419 100
FT CHAIN 420 34
FT SIGNAL 420 34
FT CHAIN 421 95
FT SIGNAL 421 95
FT CHAIN 422 100
FT SIGNAL 422 100
FT CHAIN 423 34
FT SIGNAL 423 34
FT CHAIN 424 95
FT SIGNAL 424 95
FT CHAIN 425 100
FT SIGNAL 425 100
FT CHAIN 426 34
FT SIGNAL 426 34
FT CHAIN 427 95
FT SIGNAL 427 95
FT CHAIN 428 100
FT SIGNAL 428 100
FT CHAIN 429 34
FT SIGNAL 429 34
FT CHAIN 430 95
FT SIGNAL 430 95
FT CHAIN 431 100
FT SIGNAL 431 100
FT CHAIN 432 34
FT SIGNAL 432 34
FT CHAIN 433 95
FT SIGNAL 433 95
FT CHAIN 434 100
FT SIGNAL 434 100
FT CHAIN 435 34
FT SIGNAL 435 34
FT CHAIN 436 95
FT SIGNAL 436 95
FT CHAIN 437 100
FT SIGNAL 437 100
FT CHAIN 438 34
FT SIGNAL 438 34
FT CHAIN 439 95
FT SIGNAL 439 95
FT CHAIN 440 100
FT SIGNAL 440 100
FT CHAIN 441 34
FT SIGNAL 441 34
FT CHAIN 442 95
FT SIGNAL 442 95
FT CHAIN 443 100
FT SIGNAL 443 100
FT CHAIN 444 34
FT SIGNAL 444 34
FT CHAIN 445 95
FT SIGNAL 445 95
FT CHAIN 446 100
FT SIGNAL 446 100
FT CHAIN 447 34
FT SIGNAL 447 34
FT CHAIN 448 95
FT SIGNAL 448 95
FT CHAIN 449 100
FT SIGNAL 449 100
FT CHAIN 450 34
FT SIGNAL 450 34
FT CHAIN 451 95
FT SIGNAL 451 95
FT CHAIN 452 100
FT SIGNAL 452 100
FT CHAIN 453 34
FT SIGNAL 453 34
FT CHAIN 454 95
FT SIGNAL 454 95
FT CHAIN 455 100
FT SIGNAL 455 100
FT CHAIN 456 34
FT SIGNAL 456 34
FT CHAIN 457 95
FT SIGNAL 457 95
FT CHAIN 458 100
FT SIGNAL 458 100
FT CHAIN 459 34
FT SIGNAL 459 34
FT CHAIN 460 95
FT SIGNAL 460 95
FT CHAIN 461 100
FT SIGNAL 461 100
FT CHAIN 462 34
FT SIGNAL 462 34
FT CHAIN 463 95
FT SIGNAL 463

FT CONFLICT 104 104 S -> C (IN REF. 2).
 FT CONFLICT 142 142 E -> K (IN REF. 1).
 FT CONFLICT 225 225 R -> S (IN REF. 2).
 FT CONFLICT 330 331 PL -> T (IN REF. 2).
 FT CONFLICT 337 337 E -> N (IN REF. 2).
 FT CONFLICT 356 356 E -> K (IN REF. 2).
 FT CONFLICT 362 362 G -> R (IN REF. 2).
 FT CONFLICT 369 369 MISSING (IN REF. 2 AND 3).
 FT CONFLICT 374 374 D -> N (IN REF. 2).
 FT CONFLICT 393 394 RT -> TR (IN REF. 2).
 FT CONFLICT 468 469 RK -> GR (IN REF. 2).
 FT CONFLICT 512 512 L -> P (IN REF. 2).
 FT CONFLICT 585 587 GAA -> AH (IN REF. 3).
 FT CONFLICT 601 601 A -> S (IN REF. 3).
 FT CONFLICT 628 629 DP -> RS (IN REF. 2).
 FT CONFLICT 633 633 F -> L (IN REF. 2).
 FT CONFLICT 676 676 A -> P (IN REF. 3).
 FT CONFLICT 701 701 R -> P (IN REF. 3).

Query Match 8.8%; Score 660; DB 1; Length 5703;

Best Local Similarity 22.0%; Pred. No. 6,3e-18; Matches 394; Conservative 120; Mismatches 579; Indels 698; Gaps 71;

Qy 54 MECCPDFRVCCTAELSCRGCFESFERGECDCDQCKKDYKCCPDYE--SFCALV--NRP 110
 Db 2906 VECSLDFGLVC-----RMR-----QVGRF--KMCFTNYEIRFCNCGHCHP 2944
 Qy 111 TSPSSKAPPPS-----GASQITKSTTKSPKPKKKTKKYE 150
 Db 2945 STPATSTATPSSTPGTWTILTEQTAATTATTTATSTAIPTSTPGTAPP-----KYL 2998
 Qy 151 SEELTEHSVENQESS-----SSSSSSSTIKKISS-----KNSANR----- 192
 Db 2999 SQATTPATSSKAISSSSPRTATLIPULTATKATSTATSTPTIPSTLGTGTSQNRPH 3058
 Qy 193 -----ELQKKLVKONKRNKTKKPPKPP-----VVDE-----AGS 224
 Db 3059 PMATMSTHPSSTPETHSTVLTTKATTRATSMSTSPSGTWTILTELTAATTA 3118
 Qy 225 GLDNG-----DFKVTTPDSTQHN-----KVSSTPKTTAKPI 258
 Db 3119 ALPHGTBSSTPGTWTILTEPSTAVIVPTGSTAASSTRATAGTLKVLTS--TATTPVI 3177
 Qy 259 NPR-----PSLPNSDTSKETSILVNKETTVER-----KETTTNQTSDG 300
 Db 3178 SSRATPSSPGTATAPALRSTATTPTATSVTAIPSSSLGTAMTKLSQTTPTATMSTAT 3237
 Qy 301 KEKTTSAKETOSIEKTSADLAFTSKVLAKPTPKAETTK-----GPA 343
 Db 3238 PSSPEVHTSTVLTATTTTRIGSVAPSSSTPGTAHTTKVPTTTTGFTATPSSPGTA 3297
 Qy 344 LTPP-----KEPPT-----TPKE-----PASTPKEPTP 368
 Db 3298 LTPVWISTTTPTTRGSTVTPSSIPGTHATVLTATTTTAVANGSMATPSSSTQISGTP 3357
 Qy 369 TTISAPTPKEPAPTTPKAP--TPKEPAPT--KEPAPT--TPKEPAPTTPKEPA 421
 Db 3358 PSLTTATTATATGSGTIPSSSTPGTTPIPVLTATTAATPAASTSVIPSSALGTHHPV 3417
 Qy 422 PTTTSA-----PTT-----KEPAPTTPKAPPT-----PKEP 451
 Db 3418 PNTIATHTGRSLPSSPHITVPTAMTSATSGILGTHIETPGTSHPTAATGTTQBPSTP 3477
 Qy 452 APPTPKAPTTPKEPAPTTPKEPAPT--TPKEPAPTAPKAPAPTTPKEPAPT--PKEPAP 508
 Db 3478 ALSSPHSSRTTSSPPSGITTPGHTRGTSRTTATATPSKTRSTLLPSSSPTAPITTV 3537
 Qy 509 TTTKEPS----- 515
 Db 3538 TTGCEPCAMSEMLDYSYPMFGPGGDFDTYSNIRAGAVCEOPLEGECRAQAOPGVPL 3597
 Qy 516 -----PTTKEPAPTTPKAP 531

Db 3598 RELGQVVECSLDPGVLCNRRQVQKFKMCFNYEIRFCNCGHCHPSTPATSTATSPSTP 3657
 Qy 532 TTT--KEPAPTTPSAPTTPKEBSPPTTKKEPAPTTPKEPAPTTPKAPPTTPKEPAPTTP 589
 Db 3658 GTTILATLTTTATTTATTSSTGSTATSSSTQGPAGTPH--VSTATTPVIVISK--ATP 3711
 Qy 590 KEPAPTTPKPA-----TPAPKEPA-----PTTKEPAPTTPKPLP-- 626
 Db 3712 FSSPGTATAPALASTATTPTATSTPTAISSSLGTTWTRLSQTTTPMATMTATPSSSPE 3771
 Qy 627 -----TPKEPAPTTPKEPAPT-----DEEAPTPTEED-- 656
 Db 3772 TVHTSTVLTATTTATGATGAVATSSSTPGTAHTTKVPTTTTGFTVTPSSPGTARPPV 3831
 Qy 657 ---TPTPKEPAPTTPKAPANTTPKEPAP--TPKEPAPT-----TPKEPAPT--TPKE--T 704
 Db 3832 WISTTTPPTSSSTVTPSSSIPGTHHTPVLTITTPQPAVAGSMATPSSSTQISGTPSLIT 3891
 Qy 705 APPTPKGAPTTLKEPAPTTPKAPKEPAPT--TKKEPTS-----TTSKAPAPT 753
 Db 3892 TATTITATGTT--NPSSTPGTTPIPBELTTATTPAATSTVTPSSALGTHHPVNT 3949
 Qy 754 PKGT-----APTTP-----KEPAPTTPKEPAPTTPKAPAPT--TLK 787
 Db 3950 TATTHTGRSLSPSSPHITVTAWTATSGTLGTHIETPGTSHPTAATGTTTSTPALS 4009
 Qy 788 EPAP--TPPKAPKELAPTTTGGPTSTSDKAPPTPKETAPT--PKP----- 834
 Db 4010 SPHSSRTTESPPSP--GTTTPGHTTATSRITATATPSKTRSTLLPQPTAPITTV 4065
 Qy 835 ----- 834
 Db 4066 VTGCEPCAMSEMLDYSYPMFGPGGDFDTYSNIRAGAVCEOPLEGECRAQAOPGVPL 4125
 Qy 835 -----APTTPKAPAPTTPETP 850
 Db 4126 LGEIGQVVECSLDPGVLCNRRQVQKFKMCFNYEIRFCNCGHCHPSTPATSTATSPSTP 4185
 Qy 851 PPTT--SEVSTPTTKEPTTIHKSPPDSTPELSAETPKALENSPKEPGVPTTK-- 902
 Db 4186 PGTWTILTELTATTTATTSSTGSTATP--SSTP--GTAPPKLSTPATTPATSSKATSSS 4242
 Qy 903 -----TPAATK-----PEMTTAKDKTTERDLRTTPETT 931
 Db 4243 SPRTATLIPULTSTATATSTATSVTPIPSSSLGTTGTPPEQTTTVATMTSTHPSSTPETH 4302
 Qy 932 -TAAPKMTKETATTEKT-----TESKITATTTQVTS--TTQDTPP----- 970
 Db 4303 HTSTVLTATKATTRATSTSPSSSTPGTWTILTELTAATTAGTAPATPSTPGTWTIL 4362
 Qy 971 FKITTLKTTTLAPKVTTT--KTTITTEIANKPEETAK-----PKDATNSKATTPPKQ- 1022
 Db 4363 TELTTTATTTAASSTGSTATSSSTPGTWTILTEPSTAVIVPTGSTAATSSQATAGTAPHV 4422
 Qy 1023 -----KPTKAPKKTSTKPKKT--MPRVKPKTTPPRKMT----- 1056
 Db 4423 STTATTPVSSKATTPSSSPGTATAPALASTATTPTATSTPTAISSSLGTTWTRLSQTT 4462
 Qy 1057 -----STWPELNPTRIAEAMLOTT-----TRNQTPNSKLVEVNPKESDAG- 1098
 Db 4483 TPTATMTATPSSSPPEVHTSTVLTATTTATGATGAVATSSSPGTATHTTKVPPTTTTG 4542
 Qy 1099 -GAGETPHMLLRHVMEVTPMDVLPRVPOGIIINPMLSENTNICKPKPVDGLTTL 1157
 Db 4543 TATPSSSPGATLPVWISTTTPTTTP--TTSGATVTPSSSIPGTHHTARVLTITTTTV 4600
 Qy 1158 RNLGLVAFRGHYFWMLSPPSPSPARITEVWGIPSIDIVFRCNCEGKT 1208
 Db 4601 ATGSM-----ATPSSS--TQISGTPSLTITATTTATGST 4634

RESULT 8
 PCLO_HUMAN

ID PCLO_HUMAN STANDARD; PRT; 5147 AA.
 AC Q9Y6V0; Q04373; Q06305; Q9BVC8; Q9UIV2; Q9Y6U9;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Piccolo protein (Acronin) (Fragment(s)).
 GN PCLO OR AC2 OR KIAA0559.
 OS Homo sapiens (Human).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 CC NCBI_TaxId=9606;
 CC (1)
 CC SEQUENCE OF 1-759 FROM N.A.
 CC TISSUE=Brain;
 CC MEDLINE=99439764; PubMed=10508862;
 CC Wang X., Kibbechull M., Laue M.M., Lichte B., Petrasch-Parwez E.,
 CC Kitzmann M.W.;
 CC "Acronin, a 550-kd putative scaffolding protein of presynaptic active
 CC zones, shares homology regions with rim and bassoon and binds
 CC profilin.";
 CC J. Cell Biol. 147:151-162(1999).
 CC (2)
 CC SEQUENCE OF 552-4404 FROM N.A.
 CC Kraemer J., Woliam C., Wohldmann P., McGrane B.;
 CC Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
 CC (3)
 CC SEQUENCE OF 3619-5147 FROM N.A. (ISOPFORM 2).
 CC TISSUE=Brain;
 CC MEDLINE=98290545; PubMed=9628591;
 CC Nagase T., Ishikawa K.-I., Miyajima N., Tanaka A., Kotani H.,
 CC Nomura N., Ohara O.;
 CC "Prediction of the coding sequences of unidentified human genes. IX.
 CC The complete sequences of 100 new cDNA clones from brain which can
 CC code for large proteins in vitro.";
 CC DNA Res. 5:31-39(1998).
 CC (4)
 CC SEQUENCE OF 4405-4439 FROM N.A.
 CC TISSUE=Placenta;
 CC MEDLINE=22389257; PubMed=12477932;
 CC Struhsberg R.L., Felingold E.A., Grouse L.H., Derge J.G.,
 CC Klausner R.D., Collins F.S., Wagner L., Shemen C.F., Schuler G.D.,
 CC Altschul S.F., Zeeberg B., Burow K.H., Scheffer C.F., Bhat N.K.,
 CC Hopline R.P., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 CC Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 CC Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E.,
 CC Brownstein W.J., Ueda T.B., Toshiyuki S., Carninci P., Prange C.,
 CC Raha S.S., Loggellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 CC Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 CC Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 CC Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 CC Pahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
 CC Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 CC Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 CC Rodriguez A.C., Gilwood J., Schmutz J., Myers R.M.,
 CC Butlerfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,
 CC Scherch A., Schein J.B., Jones S.J.M., Marra M.A.;
 CC "Generation and initial analysis of more than 15,000 full-length
 CC human and mouse cDNA sequences.";
 CC Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 CC (5)
 CC SEQUENCE OF 4405-5147 FROM N.A.
 CC Kalicki J., Billet G.;
 CC Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
 CC (1) FUNCTION: May act as a scaffolding protein involved in the
 CC organization of synaptic active zones and in synaptic vesicle
 CC trafficking (By similarity).
 CC (2) SUBUNIT: Interacts with Rabac1 and profilin (By similarity).
 CC (3) SUBCELLULAR LOCATION: Concentrated at the presynaptic side of
 CC synaptic junctions (By similarity).
 CC (4) ALTERNATIVE PRODUCTS:
 CC Event-Alternative splicing; Named isoforms=2;
 CC Comment=Additional isoforms seem to exist;
 CC Name=1;

CC IsoId=Q9Y6V0-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=Q9Y6V0-2; Sequence=VSP_003923, VSP_003924, VSP_003925,
 CC VSP_003926, VSP_003927;
 CC Note=No experimental confirmation available;
 CC -1- DOMAIN: C2 domain 1 is involved in binding calcium and
 CC phospholipids. Calcium binds with low affinity but with high
 CC specificity and induces a large conformational change.
 CC -1- SIMILARITY: Contains 2 C2 domains.
 CC -1- SIMILARITY: Contains 1 PDZ/DHR domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sdb.ch/announce/>
 CC or send an email to license@isb-sdb.ch).
 CC -----
 CC EMBL: Y19188; CAB60727.1; -;
 CC EMBL: AC004903; AAD20936.1; -;
 CC EMBL: AC004886; AAD21789.1; -;
 CC EMBL: AB011131; BAA25485.1; -;
 CC EMBL: BC001304; AAB01304.1; -;
 CC EMBL: AC004082; AAB97937.1; -;
 CC PIR: T00634; T00634.
 CC HSRP: P04410; 1A25.
 CC GeneW: HGNC:13406; PCLO.
 CC MIM: 604918; -;
 CC GO: GO:0005856; C:cytoskeleton; NAS.
 CC GO: GO:0045202; C:synaptic junction; ISS.
 CC GO: GO:0005509; F:calcium ion binding activity; ISS.
 CC GO: GO:0005544; F:calcium-dependent phospholipid binding actl. . .; ISS.
 CC GO: GO:0005522; F:profilin binding activity; ISS.
 CC GO: GO:0007010; P:cytoskeleton organization and biogenesis; ISS.
 CC GO: GO:0016080; P:synaptic vesicle targeting; ISS.
 CC InterPro: IPR000008; C2.
 CC InterPro: IPR015655; Synaptotagmin.
 CC PRINTS: PR00360; C2DOMAIN.
 CC PRINTS: PR00399; SYNAPTOTAGMN.
 CC SMART: SM00239; C2; 2.
 CC PROSITE: PS00499; C2 DOMAIN 1; 1.
 CC PROSITE: PS00004; C2 DOMAIN 2; 2.
 CC KEGG: Calcium/phospholipid-binding; zinc; Metal-binding; zinc-finger;
 CC Repeat; Alternative splicing.
 CC FT NON_TER 1 1
 CC FT DOMAIN 400 465
 CC FT ZN_FING 499 523
 CC FT ZN_FING 969 992
 CC FT NON_CONS 1010 1011
 CC FT DOMAIN 2300 2325
 CC FT DOMAIN 4391 4442
 CC FT DOMAIN 4544 4633
 CC FT DOMAIN 5031 5121
 CC FT VARSPLIC 4404 4404
 CC FT VARSPLIC 4534 4534
 CC FT VARSPLIC 4576 4576
 CC FT VARSPLIC 4761 4761
 CC FT VARSPLIC 4762 5147
 CC SQ SEQUENCE 5147 AA; 563537 MW; CDSB84990498CD3C CRC64;
 CC Query Match 8.8%; Score 659.5; DB 1; Length 5147;
 CC Best Local Similarity 23.7%; Pred. No. 6e-18;
 CC Matches 302; Conservative 143; Mismatches 494; Indels 337; Gaps 60;
 CC 113 PSSKKAAPPSSG-----ASQITKSTTKSPKPNKKTKVIESEITEHSVS 161

```

Db      56 PPKSGRRPDDGRAPQBLSKSRRTDTRFSGKLPGRSPSTILSKESKSRDLDEEHSS 115
Qy      162 ENOE-SSSSSSSSSSSTTMMKKS---SKNSANRELOKLVKXNDKKNRTK---KX- 212
Db      116 MPPGLSEVNLAISVSSVANKFNFPDLISDEASQEFETTKQKVQKQEGEGLIKPL 175
Qy      213 --TPKPPVDEAGSLDNGDFKVTTPDSTTQHNVSTSPKITTAKEINPPSLPBNBDT 270
Db      176 QQQPPEKLPKQGGPERDPLQODGTPKAISQOPEKIKSQPP-GTCKPIQ-GTGTPTQTDH 233
Qy      271 SKETSLTVNKEETVETKETTTTNNKQTSDEKETTSAKETOSIEKTSADLAPTSKVLAK 330
Db      234 AK---LPLQRD-----ASRPQTQADIVRESVYKPLSPSPX 267
Qy      331 PTPKAETTKGPAITTPKEPTTPPTTKEBAPTTPKEPTT---IKSAPTTPKPA---PTT 386
Db      268 PPIQOPTGPKPPAPQOPGHEKSGOPGPAK---PAQPSGLTTPKLAQOPGTIVRPPVOPGTT 323
Qy      387 KSAPTTPKEBAPTTPKEBAPTTPKBPATTTKBPATTTKBPATTPKBPATTPKBPAPT 446
Db      324 K-PPAOPIGPA---KPPAQQTGE--KPSBQEPFKLAPPPGVGKTPA---QQPQPA 372
Qy      447 TPKEBAPTTPK---EPPTTP--KEBAPTTPKEBAPTTPKEBAPTTPKBPATTPKEBAPT 501
Db      373 KPPQOVGTTPKPLAQOPGLQSPAKAPG-TKTPAQTKPSPQOPSGTKPP---PQOPGPA 427
Qy      502 TPKEBAPTTPKPS---PTTPKEBAPTTPKSAPTTPKEBAPTTPKSAPTTPKEBAPT 557
Db      428 KPSPOQPGSTPPSPSQOPGSAKPSAQOPSPAPKPSAQOPTKPVSGTGFGRPLPPTVSPSAK 487
Qy      558 KEBAPTTPKEBAP---TT-----PKK-----PAPTTPKE--- 583
Db      488 QPSOGELPKTICPLNTTELLHVEPEKANENTCIECQTVCSLGFENPHLIEAKELC 547
Qy      584 -----PAPTTPK---EPAPTTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTP 626
Db      548 LNCMKRALGGDLIAVPSSPQPKLTAVTTTSAVSKSSPQOPQSPKKA--APKQDLS 605
Qy      627 TTPKEBAPTTPKEBAPT-----TPRELAPTTPKEBAPT-----PTTPKEBAPTTPKAAAP- 674
Db      606 KAPF-----PKPEPPLVQOPPLNGSPSAKAKQPEADSLSKPAPKPSVPSSEDDKAPV 659
Qy      675 --NTPKEB--APT--PKBAPTTPKEBAPT-----PTTPKEBAPTTPKGTATTT 716
Db      660 ADDPKQPKMYKPTTDLVSSSSATTKPDIPSSKVOSQABEKTTPPLKTDASAPSSQSPPT 719
Qy      717 -----LKEBAPTTPKPAPELAPTTPKEPTSTSDKPAV--- 751
Db      720 GEKVTTPDSKAIIPRASDSKIIISHGPSSESK-GQKQVDPVQKKEPKKQTKMSPKDA 778
Qy      752 -TTPKGTAPTTPKEBAPTTPKEBAPTTPKGTAP-----TTLKEBAPTTPK 796
Db      779 KPMFGS---PTPPGPRPTA-GQIVPTPOQSPKPOQSRFRFSMLTGSIDAVKSGQPTTPOE 835
Qy      797 PAPKEL-----APTTPKGSTSTSDKPAPELAPTTPKEBAPTTPKBPAP- 834
Db      836 TVTGKLFEGFASISQASNLISTAGQPGPHSOSGGAQPMQAIPASQP---PTSGPFPK 891
Qy      835 -----APTTPKBPAPTTPKEPTTPSEV-----STPTTPKEPTTTHKSDPSTELASBP- 884
Db      892 STGQAPAPAKSIPVKEKTKAPAAEKLEPKABQAPTVARTETKEKPPRIKOSKSLTAPQ 951
Qy      885 ---TPKALENSPK-EPGVPTTKTP---AATPEKTTAKDKT----- 919
Db      952 KAVLPTKLEKSPKESTCPLCTEINISKDPFNNTCTEKNQVNCNGFPTPHLTEN 1011
Qy      920 ---TER-----DRTPEETTA--APMTKETATTEKTT---ESKITATTTQVTSSTT 965
Db      1012 CQTOAISGQGLDKRMPPADSGPAPSPVPTSSSQKTAVPPVOKLVKQEQGVKTEA 1071
Qy      966 QDTTPFKI-TTLKTTTLAPKVTTTKTITTEINMKPEETAKPKORATNSKATTPKPKP 1024

```

```

Db      1072 EKVLIEKVEKTELMEKILPMVTTDQ-----KQESKLEKDKASALQEKKLPBEEK 1121
Qy      1025 TKAPKCTSTKPKKPTMVRVAKPKTTPPRK-----TSTME-----LNPFSRIAEAMQ 1074
Db      1122 KLIPEEKIRISEEK-KPLIEKPKTPEKDKLIPAKTSABEEOGHDLKQVQIAEKELE 1180
Qy      1075 TTPRNPQTPNSKIVENVKPSDAGAGETPPHMLRPHVFMPEVTPPMDYLPRVNOGII 1134
Db      1181 GRVAPKTVQBEK-----QPTKMEGLPSG-TPQSLPKD---DKTKTKIKQPPQPP----- 1227
Qy      1135 INPMLSDETINCSKP 1150
Db      1228 -----CTAKP 1232

RESULT 9
MUC1 HUMAN
ID MUC1 HUMAN STANDARD; PRT; 1255 AA.
AC P15941; P13931; P15942; P17626; Q14128; Q14876; Q16437; Q16442;
AC Q16615; Q9BX44; Q9UB75; Q9UB76; Q9UOL1; Q9Y4J2;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Mucin 1 precursor (MUC-1) (Polymorphic epithelial mucin) (PEM) (PEMT)
DE (Epistatin) (Tumor-associated mucin) (Carcinoma-associated mucin)
DE (Tumor-associated epithelial membrane antigen) (EMA) (H2JAG) (Peanut-
DE reactive urinary mucin) (PUM) (Breast carcinoma-associated antigen
DE DF3) (CD227 antigen).
GN MUC1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP TISSUE=PANCREAS;
RC MUC1IN=90368716; PubMed=2394722;
RA Lan M.S., Batria S.K., Qi W.-N., Metzgar R.S., Hollingsworth M.A.;
RT "Cloning and sequencing of a human pancreatic tumor mucin cDNA.";
RL J. Biol. Chem. 265:15294-15299(1990).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RA MEDLINE=90202794; PubMed=2318825;
RT Lichtenberg M.J.L., Vos H.L., Gennissen A.M.C., Hilkens J.;
RT "Epistatin, a carcinoma-associated mucin, is generated by a
RT polymorphic gene encoding splice variants with alternative amino
RT termini.";
RL J. Biol. Chem. 265:5573-5578(1990).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Breast carcinoma;
RA MEDLINE=90368715; PubMed=1697589;
RT Gendler S.J., Lancaster C.A., Taylor-Papadimitriou J., Duhig T.,
RT Peat N., Burchell J., Pemberton L., Lalani E.-N., Wilson D.;
RT "Molecular cloning and expression of human tumor-associated
RT polymorphic epithelial mucin.";
RL J. Biol. Chem. 265:15286-15293(1990).
RN [4]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RA MEDLINE=91097524; PubMed=2268309;
RT Lancaster C.A., Peat N., Duhig T., Wilson D.,
RT Taylor-Papadimitriou J., Gendler S.J.;
RT "Structure and expression of the human polymorphic epithelial mucin
RT gene: an expressed VNTR unit.";
RL Biochem. Biophys. Res. Commun. 173:1019-1029(1990).
RN [5]
RP SEQUENCE FROM N.A. (ISOFORM 5).
RC TISSUE=Breast carcinoma;
RA MEDLINE=90276413; PubMed=2351132;
RT Wareschneider D.H., Harevuani M., Tsefaty I., Smorodinsky N., Horev J.,
RT Zaretsky J., Kocbes P., Weiss M., Lache R., Dion A., Keydar I.;
RT "Human epithelial tumor antigen cDNA sequences. Differential splicing
RT may generate multiple protein forms.";

```

RL Eur. J. Biochem. 189:463-473(1990).
 RN [6]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Breast carcinoma;
 RX MEDLINE=90276414; PubMed2112460;
 RA Harevent M., Tsafaty I., Zaretsky J., Kotkes P., Horev J.,
 RT Zrlan S., Weiss M., Green S., Lach R., Keydar I., Wreschner D.H.;
 RT "A transcribed gene, containing a variable number of tandem repeats,
 RT codes for a human epithelial tumor antigen. cDNA cloning, expression
 RT of the transcribed gene and over-expression in breast cancer
 RT tissue.";
 RL Eur. J. Biochem. 189:475-486(1990).
 RN [7]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RC MEDLINE=9103045; PubMed1688329;
 RX Tsafaty I., Harevent M., Horev J., Zaretsky J., Weiss M.,
 RA Teitelbaum J., Garnier J.M., Lach R., Keydar I., Wreschner D.H.;
 RT "Isolation and characterization of an expressed hypervariable gene
 RT coding for a breast-cancer-associated antigen.";
 RL Gene 93:313-318(1990).
 RN [8]
 RP SEQUENCE FROM N.A. (ISOFORM 7).
 RC MEDLINE=95010060; PubMed7925397;
 RX Zrlan-Licht S., Vos H.L., Baruch A., Elroy-Stein O., Sagiv D.,
 RA Keydar I., Hilkens J., Wreschner D.H.;
 RT "Characterization and molecular cloning of a novel MUC1 protein,
 RT devoid of tandem repeats, expressed in human breast cancer tissue.";
 RL Eur. J. Biochem. 224:787-795(1994).
 RN [9]
 RP SEQUENCE FROM N.A. (ISOFORMS 6; 7 AND 8).
 RC MEDLINE=97355747; PubMed9212228;
 RX Oosterkamp H.M., Scheiner L., Stefanova M.C., Lloyd K.O.,
 RA Pineda C.L.;
 RT "Comparison of MUC-1 mucin expression in epithelial and non-epithelial
 RT cancer cell lines and demonstration of a new short variant form
 RT (MUC-1/2).";
 RL Int. J. Cancer 72:87-94(1997).
 RN [10]
 RP SEQUENCE FROM N.A. (ISOFORM 7).
 RC Zhang L.X., Li C.H.;
 RT "Molecular cloning of an isoform of MUC1, MUC1/X.";
 RL Submitted (Feb-1999) to the EMBL/GenBank/DBJ databases.
 RN [11]
 RP SEQUENCE FROM N.A. (ISOFORM 9).
 RC TISSUE-Epithelial cancer;
 RX Zhang L.X., Li C.H., Sun L.Y., Yue W.;
 RT "Cloning of a new potential secreted short variant form of MUC1 mucin
 RT in epithelial cancer cell line.";
 RL Submitted (Feb-2001) to the EMBL/GenBank/DBJ databases.
 RN [12]
 RP PARTIAL SEQUENCE FROM N.A.
 RC MEDLINE=88330762; PubMed3417635;
 RX Gendler S.J., Taylor-Papadimitriou J., Duhig T., Rotherhard J.,
 RA Burchell J.;
 RT "A highly immunogenic region of a human polymorphic epithelial mucin
 RT expressed by carcinomas is made up of tandem repeats.";
 RL J. Biol. Chem. 263:12820-12823(1988).
 RN [13]
 RP SEQUENCE OF 1-160 FROM N.A. (ISOFORM 2).
 RC MEDLINE=9008473; PubMed2597151;
 RX Abe M., Siddiqui J., Kufe D.;
 RT "Sequence analysis of the 5' region of the human DP3 breast
 RT carcinoma-associated antigen gene.";
 RL Biochem. Biophys. Res. Commun. 165:644-649(1989).
 RN [14]
 RP SEQUENCE OF 1-109 FROM N.A. (ISOFORM 2).
 RC TISSUE=Thyroid;
 RX MEDLINE=96183746; PubMed8608966;
 RA Weiss M., Baruch A., Keydar I., Wreschner D.H.;
 RT "Preoperative diagnosis of thyroid papillary carcinoma by reverse
 RT transcriptase polymerase chain reaction of the MUC1 gene.";
 RL Int. J. Cancer 66:55-59(1996).
 RN [15]

RP SEQUENCE OF 1-89 FROM N.A.
 RC TISSUE=Lung;
 RX MEDLINE=96181716; PubMed8604237;
 RA Yu C.J., Yang P.C., Shew J.Y., Hong T.M., Yang S.C., Lee Y.C.,
 RT Lee L.N., Luh K.T., Wu C.W.;
 RT "Mucin mRNA expression in lung adenocarcinoma cell lines and
 RT tissues.";
 RL Oncology 53:118-126(1996).
 RN [16]
 RP SEQUENCE OF 1-46 FROM N.A. (ISOFORMS 3 AND 4).
 RC TISSUE-Breast carcinoma;
 RX Bulwela L., Liu Q., Jugmani Y.A., Gomm J.J., Coombes R.C.;
 RL Submitted (Oct-1992) to the EMBL/GenBank/DBJ databases.
 RN [17]
 RP CARBOHYDRATE-LINKAGE SITES WITHIN THE REPEAT.
 RX MEDLINE=97460054; PubMed9312074;
 RA Mueller S., Goletz S., Packer N., Gooley A.A., Lawson A.M.,
 RA Hantsch F.-G.;
 RT "Localization of O-glycosylation sites on glycopeptide fragments from
 RT lactation-associated MUC1. All putative sites within the tandem
 RT repeat are glycosylation targets in vivo.";
 RL J. Biol. Chem. 272:24780-24793(1997).
 RN [18]
 RP CARBOHYDRATE-LINKAGE SITES WITHIN THE REPEAT.
 RX MEDLINE=99303572; PubMed10373415;
 RA Mueller S., Alving K., Peter-Katalinic J., Zachara N., Gooley A.A.,
 RA Hantsch F.-G.;
 RT "High density O-glycosylation on tandem repeat peptide from secretory
 RT MUC1 of T47D breast cancer cells.";
 RL J. Biol. Chem. 274:18165-18172(1999).
 RN [19]
 RP POLYMORPHISM WITHIN THE REPEAT.
 RX MEDLINE=21359366; PubMed11350974;
 RA Engelmann K., Baldu S.B., Hantsch F.-G.;
 RT "Identification and topology of variant sequences within individual
 RT repeat domains of the human epithelial tumor mucin MUC1.";
 RL J. Biol. Chem. 276:27764-27769(2001).
 RN [20]
 RP CHARACTERIZATION OF ISOFORM Y, AND MUTAGENESIS OF ASP-1116.
 RX MEDLINE=99211485; PubMed10197628;
 RA Baruch A., Hartmann M.-L., Yoeli M., Adereth Y., Greenstein S.,
 RA Stadler Y., Skornik Y., Zaretsky J., Smorodinsky N.I., Keydar I.,
 RA Wreschner D.H.;
 RT "The breast cancer-associated MUC1 gene generates both a receptor and
 RT its cognate binding protein.";
 RL Cancer Res. 59:1552-1561(1999).
 RN [21]
 RP PARTIAL SEQUENCE, AND CHARACTERIZATION OF CLEAVAGE SITE.
 RX MEDLINE=21240104; PubMed11341784;
 RA Parry S., Silverman H.S., McDermott K., Willis A., Hollingsworth M.A.,
 RA Harris A.;
 RT "Identification of MUC1 proteolytic cleavage sites in vivo.";
 RL Biochem. Biophys. Res. Commun. 283:715-720(2001).
 RN [22]
 RP CHARACTERIZATION.
 RX MEDLINE=21836452; PubMed11847293;
 RA Wreschner D.H., McGuckin M.A., Williams S.J., Baruch A., Yoeli M.,
 RA Ziv R., Okun L., Zaretsky J., Smorodinsky N., Keydar I., Neophytou P.,
 RA Stacey M., Lin H.-H., Gordon S.;
 RT "Generation of ligand-receptor alliances by 'SEA' module-mediated
 RT cleavage of membrane-associated mucin proteins.";
 RL Protein Sci. 11:698-706(2002).
 RN [23]
 RP PHOSPHORYLATION.
 RX MEDLINE=95080414; PubMed7988707;
 RA Zrlan-Licht S., Baruch A., Elroy-Stein O., Keydar I., Wreschner D.H.;
 RT "Tyrosine phosphorylation of the MUC1 breast cancer membrane proteins.
 RT Cytokine receptor-like molecules.";
 RL FEBS Lett. 356:130-136(1994).
 CC -1- FUNCTION: May play a role in adhesive functions and in cell-cell
 CC interactions, metastasis and signaling. May provide a protective

Query Match

8.7%; Score 651; DB 1; Length 1255;

Best Local Similarity 27.9%; Pred. No. 3.9e-18;
Matches 294; Conservative 74; Mismatches 466; Indels 216; Gaps 50;

```

Qy 274 TSLVNNKETTVEKETTNNKQSTDGKETSIAKETSIEKTSKD-LAPTSKYLAAPT 332
Db 16 TLTIV-----VTSGHASTPGGEKETSATQSRVSPSSTKKNVSMTSVLSHS 65
Qy 333 P-KAETTTKGP--ALTTPKEP-----TPTTPKEPASTTKEPTPTI---K 372
Db 66 PGSSSTTQGDVTLAPATEPSSAATWGODVTSVPTRPALGSTTPAHVDVSAIPNK 125
Qy 373 SAPTPKEPAPTTTKSAPTTKEPAPTTTKEP-----APTTKEPAPTTTKEPAPTTK 426
Db 126 PAPGSTAPPAHGVTSAPDT--RPAHGSTAPPAHGVTSAPDT--RPAHGSTAPPAHGVTS 179
Qy 427 SAPTPKEPAPTTK---KPAPTTKEPAPTTKEPPTTKEP--PAPTTKEP----- 474
Db 180 SAPTPRAPAGSTAPPAHGVTSAPDTRRAPAGSTAPPAHGVTSAPDTRRAPAGSTAPPAHGVTS 239
Qy 475 -APTPKEPAPTAAPK---KPAPTTKEPAPTTK---EPAPTTTKEPSPPTTKEPAPT 525
Db 240 SAPDTRRAPAGSTAPPAHGVTSAPDTRRAPAGSTAPPAHGVTSAPDTRRAPAGSTAPPAHGVTS 299
Qy 526 T---TKSAPTTTKEPAPTTTKSAPTTKEPSPPTTKEP-----APTTKEPAPTTK-- 574
Db 300 SAPDTRRAPAGSTAPPAHGVTSAPDT--RPAHGSTAPPAHGVTSAPDTRRAPAGSTAPPAH 356
Qy 575 --KPAPTTKEPAPTTK---EPAPTTTKEPAPTAAPK---KPAPTTKEPAPTTK 628
Db 357 GVTSAAPDTRRAPAGSTAPPAHGVTSAPDTRRAPAGSTAPPAHGVTSAPDTRRAPAGSTAPPAH 416
Qy 629 PEKLAAPTEPEKAPPTPEELAPTTPEEPTPTTPEEAPATTKA-----AAPPTKEPAPT 684
Db 417 GVTSAAPDTRRAPAGSTAP--PAHGVTSAPDTRRAPAGSTAPPAHGVTSAPDTRRAPAGSTA 472
Qy 685 PK---BEAPTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT-----KKPAP 730
Db 473 PPAHGVTSAPDTRRAPAGSTAP--PAHGVTSAPDTRRAPAGSTAPPAHGVTSAPDTRRAP 528
Qy 731 KELAPTTTKEPTSTSDKAP--TTPKGTAPTTTKEPAPTTTKEPAPTTKPG---TAPT 784
Db 529 GSTAP--PAHGVTSAPDTRRAPAGSTAPPAHGVTS-----APDTRRAPAGSTAPPAHGVTSAPD 583
Qy 785 TLKEPAPPTP-----KKPAPKELAPTTTGSTSTSDKAP--TTPK---ETAP 828
Db 584 TRAPAGSTAPPAHGVTSAPDTRRAPAGSTAP--PAHGVTSAPDTRRAPAGSTAPPAHGVTSAP 642
Qy 829 TTPKEPAPPTP-----KKPAPTTPEPTPTTSEVSTPTTKEPTT-----IHK 871
Db 643 DTRAPAGSTAPPAHGVTSAPDTRRAPAGS--TAPPAHGVTSAPDTRRAPAGSTAPPAHGVTS 700
Qy 872 SPDSSTPELSAEPPTKALENSPEKSGVTTTTPA-----TKPEMTTAKDK--TT 920
Db 701 APDTRRAPAGSTAPPAHGVTSAPDTRRAPAGSTAPPAHGVTSAPDTRRAPAGSTAPPAHGVTS 760
Qy 921 ERDRTTPEPT-----TAAPKMTKEATTTTEKTESKITATTTQVSTTQDTP--F 971
Db 761 APDTRRAPAGSTAPPAHGVTSAP--DTRRAPAGSTAPPAHGVTSAPDTRRAPAGSTAPPAH 816
Qy 972 KITTLKTTTTLAPKVTTTKKTTTT--EIMNPEETAKEDRATNSKATTPKOKPTKAPK 1029
Db 817 GVTSAAPDTRRAPAGSTAPPAHGVTSAPDTRRAPAGSTAPPAHGVTSAPDTRRAP--GSTAP 874
Qy 1030 KPSTTKKPKTMRPRKPTTPTPRKMTSTMELNPTSIAMAMQTTTRPQCT--PNSKLV 1088
Db 875 AHGVTSAPDTRP--APGSTAPPAHGVTSAPDTRPA-----PGSTAPPAHGV 918
Qy 1089 EVNPKSDAGGAEGETPMLRPHVMEVTPMDVYLPRVNOGIIINPMKSDFTNIG 1148
Db 919 TSADTTRAPAGSTAPPAH-----GVTSAPNRRALDSTA---PVAHVNTSASGS 964
Qy 1149 KPVVDGLTTLRNGTLVAFRGHYFMMLSPPSPS 1180

```

```

Db 965 ASGSASTLVHNGTSARATTTTPASKSTPSPSIPS 996

RESULT 10
UN89 CAEEL STANDARD; PRT; 6632 AA.
AC 001761; Q17362;
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DE Muscle M-line assembly protein unc-89 (Uncoordinated protein 89).
GN UNC-89 OR C09D1.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A., FUNCTION, AND TISSUE SPECIFICITY.
RC STRAIN=Bristol N2;
RX MEDLINE=96180278; PubMed=8603916;
RA Benham G.M., Tinley T.L., Tang X., Borodovsky M.;
RT "The Caenorhabditis elegans gene unc-89, required for muscle M-line
RT transduction domains."
RL J. Cell Biol. 132:835-848(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Du Z., Le T.T., Wilson R.;
RN Submitted (May-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP REVISIONS.
RA Waterston R.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Structural component of the muscle M-line. Myofibril
CC lattice assembly begins with positional cues laid down in the
CC basement membrane and muscle cell membrane. UNC-89 responds to
CC these signals, localizes, and then participates in assembling an
CC M-line.
CC -1- TISSUE SPECIFICITY: Localizes to the middle of A-bands.
CC -1- SIMILARITY: Contains 1 DBJ-homology (DH) domain.
CC -1- SIMILARITY: Contains 1 fibronectin type III domain.
CC -1- SIMILARITY: Contains 49 immunoglobulin-like C2-type domains.
CC -1- SIMILARITY: Contains 1 PH domain.
CC -1- SIMILARITY: Contains 5 RGD domains.
CC -1- SIMILARITY: Contains 1 SH3 domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U33058; AAB00542.1; -
CC EMBL: AF003131; AAB54132.2; -
CC PDB: 1FHO; 20-DEC-00.
CC Wormpep; C09D1.1; CE30426.
CC InterPro: IPR003961; FN III.
CC InterPro: IPR007110; Ig-like.
CC InterPro: IPR003598; Ig C2.
CC InterPro: IPR003006; Ig_MHC.
CC InterPro: IPR001849; PH.
CC InterPro: IPR007850; RGD.
CC InterPro: IPR000219; RHOGEF.
CC InterPro: IPR001452; SH3.
CC Pfam: PF000041; fn3; 1.
CC Pfam: PF00047; ig; 47.
CC Pfam: PF00169; PH; 1.
CC Pfam: PF05177; RGD; 5.
CC Pfam: PF00621; RHOGEF; 1.
CC Pfam: PF00018; SH3; 1.

```

DR SMART; SM00408; IgG2; 23.
DR SMART; SM00325; RhogE; 1.
DR PROSITE; PS00326; SH3; 1.
DR PROSITE; PS50010; DH; 2; 1.
DR PROSITE; PS50835; IG_LIKE; 49.
DR PROSITE; PS50003; PH_DOMAIN; 1.
DR PROSITE; PS50002; SH3; 1.
KW Muscle Protein; Immunoglobulin domain; Repeat; SH3 domain;
3D-structure.
FT DOMAIN 63 127 SH3.
FT DOMAIN 152 330 DH.
FT DOMAIN 342 498 PH.
FT DOMAIN 547 633 IG-LIKE C2-TYPE 1.
FT DOMAIN 736 748 IG-LIKE C2-TYPE 2.
FT DOMAIN 748 838 IG-LIKE C2-TYPE 3.
FT DOMAIN 946 1033 IG-LIKE C2-TYPE 4.
FT DOMAIN 1044 1132 IG-LIKE C2-TYPE 5.
FT DOMAIN 1140 1227 IG-LIKE C2-TYPE 6.
FT DOMAIN 1272 1315 THR-RICH.
FT DOMAIN 1375 1475 RCSD 1.
FT DOMAIN 1479 1585 RCSD 2.
FT DOMAIN 1597 1695 RCSD 3.
FT DOMAIN 1700 1799 RCSD 4.
FT DOMAIN 1800 1860 RCSD 5.
FT DOMAIN 1982 2067 IG-LIKE C2-TYPE 7.
FT DOMAIN 2071 2163 IG-LIKE C2-TYPE 8.
FT DOMAIN 2171 2261 IG-LIKE C2-TYPE 9.
FT DOMAIN 2269 2359 IG-LIKE C2-TYPE 10.
FT DOMAIN 2367 2455 IG-LIKE C2-TYPE 11.
FT DOMAIN 2463 2564 IG-LIKE C2-TYPE 12.
FT DOMAIN 2563 2651 IG-LIKE C2-TYPE 13.
FT DOMAIN 2657 2746 IG-LIKE C2-TYPE 14.
FT DOMAIN 2754 2858 IG-LIKE C2-TYPE 15.
FT DOMAIN 2887 2980 IG-LIKE C2-TYPE 16.
FT DOMAIN 2994 3081 IG-LIKE C2-TYPE 17.
FT DOMAIN 3087 3183 IG-LIKE C2-TYPE 18.
FT DOMAIN 3189 3280 IG-LIKE C2-TYPE 19.
FT DOMAIN 3286 3376 IG-LIKE C2-TYPE 20.
FT DOMAIN 3384 3469 IG-LIKE C2-TYPE 21.
FT DOMAIN 3482 3572 IG-LIKE C2-TYPE 22.
FT DOMAIN 3580 3667 IG-LIKE C2-TYPE 23.
FT DOMAIN 3686 3777 IG-LIKE C2-TYPE 24.
FT DOMAIN 3817 3908 IG-LIKE C2-TYPE 25.
FT DOMAIN 3920 4009 IG-LIKE C2-TYPE 26.
FT DOMAIN 4018 4106 IG-LIKE C2-TYPE 27.
FT DOMAIN 4201 4201 IG-LIKE C2-TYPE 28.
FT DOMAIN 4212 4297 IG-LIKE C2-TYPE 29.
FT DOMAIN 4302 4387 IG-LIKE C2-TYPE 30.
FT DOMAIN 4400 4485 IG-LIKE C2-TYPE 31.
FT DOMAIN 4489 4580 IG-LIKE C2-TYPE 32.
FT DOMAIN 4588 4678 IG-LIKE C2-TYPE 33.
FT DOMAIN 4681 4771 IG-LIKE C2-TYPE 34.
FT DOMAIN 4873 4961 IG-LIKE C2-TYPE 35.
FT DOMAIN 4965 5057 IG-LIKE C2-TYPE 36.
FT DOMAIN 5067 5160 IG-LIKE C2-TYPE 37.
FT DOMAIN 5171 5260 IG-LIKE C2-TYPE 38.
FT DOMAIN 5277 5366 IG-LIKE C2-TYPE 39.
FT DOMAIN 5383 5472 IG-LIKE C2-TYPE 40.
FT DOMAIN 5487 5578 IG-LIKE C2-TYPE 41.
FT DOMAIN 5595 5685 IG-LIKE C2-TYPE 42.
FT DOMAIN 5701 5790 IG-LIKE C2-TYPE 43.
FT DOMAIN 5815 5904 IG-LIKE C2-TYPE 44.
FT DOMAIN 5925 6014 IG-LIKE C2-TYPE 45.
FT DOMAIN 6038 6130 IG-LIKE C2-TYPE 46.
FT DOMAIN 6150 6239 IG-LIKE C2-TYPE 47.
FT DOMAIN 6275 6368 FIBROBLASTIN TYPE-II.
FT DOMAIN 6413 6502 IG-LIKE C2-TYPE 48.
FT DOMAIN 6507 6596 IG-LIKE C2-TYPE 49.
FT DISULFID 568 621 POTENTIAL.
FT DISULFID 2908 2975 POTENTIAL.
FT DISULFID 3015 3065 POTENTIAL.
FT DISULFID 3707 3759 POTENTIAL.
FT DISULFID 3826 3890 POTENTIAL.

FT DISULFID 5092 5157 POTENTIAL.
FT DISULFID 5298 5350 POTENTIAL.
FT DISULFID 5508 5560 POTENTIAL.
FT DISULFID 5616 5669 POTENTIAL.
FT DISULFID 5722 5764 POTENTIAL.
FT DISULFID 5836 5901 POTENTIAL.
FT DISULFID 5946 5998 POTENTIAL.
FT DISULFID 6036 6171 POTENTIAL.
FT DISULFID 6421 6486 POTENTIAL.
FT CONFICT 2137 2137 A -> P (IN REF. 1).
FT CONFICT 2245 2247 ACA -> PKP (IN REF. 1).
FT CONFICT 2258 2258 A -> P (IN REF. 1).
FT CONFICT 2284 2284 E -> G (IN REF. 1).
FT CONFICT 2297 2297 M -> I (IN REF. 1).
FT CONFICT 3531 3531 A -> G (IN REF. 1).
FT CONFICT 3531 3531 DAGEY -> RRRRI (IN REF. 1).
FT CONFICT 3864 3868 A -> V (IN REF. 1).
FT CONFICT 3929 3929 A -> P (IN REF. 1).
FT CONFICT 5134 5134 T -> S (IN REF. 1).
FT CONFICT 5145 5145 G -> A (IN REF. 1).
FT CONFICT 5185 5185 K -> N (IN REF. 1).
FT CONFICT 5199 5199 L -> F (IN REF. 1).
FT CONFICT 5202 5202 F -> L (IN REF. 1).
FT CONFICT 5213 5213 A -> G (IN REF. 1).
FT CONFICT 6178 6178 K -> E (IN REF. 1).
FT CONFICT 6268 6268
SQ SEQUENCE 6632 AA; 731665 MW; 262D3EDD62960E89 CRC64;
Query Match 8.4%; Score 629; DB 1; Length 6632;
Best Local Similarity 23.4%; Pred. No. 1e-16;
Matches 273; Conservative 156; Mismatches 467; Indels 270; Gaps 44;
141 NKKTKKYIESEIEHSEVSENOR-----SSSSSSSSSSSTIMIKSKSKSNAARELOK 196
1176 NKRKRVDOEGARVVRDSYVDGASILTIDATVYSEVNHLSYV--AENTLGAETGA 1233
197 KLVKVDNKNKPTKPKPVVDAGSLDNGDFKVTTPDTSYOHNK-----VSTSP 250
1234 QLTIPKESVYVER-----QDLSSSEVKEIKOQKEASP 1269
251 KITLAKPINPRPSLDPNSDTSKEISLTVNKEETVETKSTTTNNKQSTDGKEKTSANET 310
1270 EATTT-----ITWETSLSTKTTMTSTEVTSVGVGVETKSESSESATV 1315
311 Q-----SIETSAKDLPATSKVLAKPTPKKETTGGALTTTPKEPTTPPKS 357
1316 VIGGGGGVTEGSSISVSKIEVSKTSDQTVREGPKRVRGSAEEL--PKREVDSRRK 1373
358 PASTTP--KEPTPTTIKAPPTPKKEPAPPTTKSAPPTPKKEPAPPTTKKEPAPT 415
1374 KSPSPDCKKESPEKTEKPKASPTKKGEBVKS-----PKE-----KSPASTPKKESPA 1423
416 TTKEBAPTTTYSAPPTP--KEBAPTTPKKAPPTPKKEBAPTTPKKEBAPTTPKE 473
1424 AEEVSPKPKESPSPTKKEKSPSPKTKGDEVKESP-----PK--SPTKKE 1471
474 PAPTPKPPAPAPAKKAPPTPKKEBAPTTPKKEBAPTTPKESPTT--PKBAPTTKSAPT 532
1472 KSPKPEDEVKSPVKKK-----KSPDATNIVEVSEETIEKETTTMTTMEHSESESRIS 1525
533 TTKEBAPTTTYSAPPTP---KEBPTTKKEBAPPTPKKEBAP--TPKAPAPTTPKEAPT 587
1526 VKKEKTEPKVDEKPSPTKOKSPKSIITBEIKSVKKEKSPKKEBKEKPASTTKKESBP 1585
588 TPKEBAPTTTKKAPAPAKKEBAP-----TPKKEAP--TPPKLPTTPEKLAAPTTP 637
1586 KPASPTKSENVKSPKPKOKSPKSEKSVVEELKSPKESPKAKDDPKSPKPKK-----SP 1641
638 EKPA-----PTPEELAPTTPEBPTPTTPEBAPT--TPKAAANTPKBAPPTPKK--P 688
1642 EKSATEDVKSPTKKEKSPKVEKESPTKKESSPTKTDDEVKSPTKKESKSPQVEKBP 1701
689 APPTKEBAPTTPKETAPPTPKGTAPPTLKKEBAPPTPKKAPAKELAPTTKEPTSTSDK 748

```

Db 1702 ASPTKEKSPKSVVEVYKSPKSP-----EKAEEKSPKPTKE-----KSPEKSAEE 1751
Qy 749 PAPTPKGTAPTTREPAPTTPKKAPTTPGTAPTLTKBPAPTTPKKAPKAPKELAPTTK 808
Db 1752 VKSPFKKSPKSPKSAEEKPKSP-----TKKSSSVKN---ADDEVKSPTKKESKP----- 1798
Qy 809 GPTSTSDKPAPTTPKETAAPTTPKBPAPTTPKKAPTTPEPTTPPTTSEVSTPTTKKEPTT 868
Db 1799 ---EKVEKPAKSPKKE-----KTPKESAAEELKSPTKKESKSPSSPTKKTGDS 1844
Qy 869 IHKSPDESTPELSAPPTPKALENSKPEGVPTTKTPAATKCBMTTAAKTKTERDLKTP 928
Db 1845 KEKSP-EKPEEKPKSPKTPK---KSPGSPKSKKSKSPKAPK---PAKLTPLDLQTVN 1897
Qy 929 ETTAAPKMTKBTATTEKTESKLTATTVQSTTTQ-----DTPPKITTLKTT 980
Db 1898 KTDLAHFVVEVHAETCKMFLDGKEITTAQGVASKDOFEFRCSIDTTFGSGTV--SV 1955
Qy 981 LAPKVTTKKTTTTEIMNKPEETAKP----- 1007
Db 1956 VASNAAGSVETKELKVLTPRETCKKPPFTDKLRMEVTKGDTVMQVIALHSPLYKMYQ 2015
Qy 1008 -----KDRATNSKATTPKPO---KPTKAPKKTSTYKPKTPMPVRKPKTTP 1050
Db 2016 NGNLEDEKNGVTTIKNEENKSSLIIPNADSGKITVEASNEVGSSESAQLTVNPPSTTP 2075
Qy 1051 T-----PRKMTSTMPKL-----NPTSR-----IAEAMQTTTRN-----QTP 1083
Db 2076 IVDGPKSVTTKETETAPKATISGFPAPTVMTINEKIVESRTITTKEDVYTKIS 2135
Qy 1084 NSKIVE---VNPKSEDAAGETPHMLRPHVPEVTPDMVDYLPVYNOG-----I 1133
Db 2136 NAKIEGTGTAVTANASGAGDSKQADLKVPRVAKPKRSQI--TDKXADESEPLRMNLE 2193
Qy 1134 IINPMLEPTN-ICNGKPVDTLTLR 1158
Db 2194 LDGSPGTEVSWMLNGOPLTKSDTVQ 2219

RESULT 11
ID PCLO CHICK STANDARD; PRT; 5120 AA.
AC QEPV36;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DB Piccolo protein (Aczonin) (Fragment).
GN PCLO OR ACZ.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianine;
OC Gallus.
OC NCBI_TaxID=9031;
RX MEDLINE=99439764; PubMed=10508862;
RA Wang X., Kibechull M., Laue M.M., Lichte B., Petrasch-Parwez E.,
RA Kilimann M.W.
RT "Aczonin, a 550-kD putative scaffolding protein of presynaptic active
RT zones, shares homology regions with rim and bassoon and binds
RT profilin".
RT J. Cell Biol. 147:151-162(1999).
CC -1- FUNCTION: May act as a scaffolding protein involved in the
CC organization of synaptic active zones and in synaptic vesicle
CC trafficking (By similarity).
CC -1- SUBUNIT: Interacts with Rabac1/Pral and profilin (By similarity).
CC -1- SUBCELLULAR LOCATION: Concentrated at the presynaptic side of
CC synaptic junctions (By similarity).
CC -1- DOMAIN: C2 domain 1 is involved in binding calcium and
CC phospholipids. Calcium binds with low affinity but with high
CC specificity and induces a large conformational change.
CC -1- SIMILARITY: Contains 2 C2 domains.

```

```

CC -1- SIMILARITY: Contains 1 PDZ/DHR domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements
CC or send an email to license@ebi.ac.uk).
CC -----
DR EMBL; Y19187; CAB60725.1; -.
DR HSSP; P04410; 1A25.
DR GO; GO:0045202; C:synaptic junction; ISS.
DR GO; GO:0005509; F:calcium ion binding activity; ISS.
DR GO; GO:0005544; F:calcium-dependent phospholipid binding acti. . .; ISS.
DR GO; GO:0005522; F:profilin binding activity; ISS.
DR GO; GO:0007010; F:cytoskeleton organization and biogenesis; ISS.
DR GO; GO:0016080; P:synaptic vesicle targeting; ISS.
DR InterPro; IPR000008; C2.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001565; Synaptotagmin.
DR Pfam; PF00168; C2; 2.
DR Pfam; PF00595; PDZ; 1.
DR PRINTS; PR00399; SYNAPTOTAGMN.
DR SMART; SM00239; C2; 2.
DR SMART; SM00228; PDZ; 1.
DR PROSITE; PS00499; C2 DOMAIN 1; 1.
DR PROSITE; PS50004; C2 DOMAIN 2; 2.
DR PROSITE; PS50106; PDZ; 1.
DR Calcium/phospholipid-binding; Zinc; Metal-binding; Zinc-finger;
KW Repeat.
KW NON TER.
FT DOMAIN 1
FT 258 357
FT 10 X 10 AA TANDEM APPROXIMATE REPEATS OF
FT ZN FING 368 392 P-A-K-P-Q-P-Q-P-X.
FT ZN FING 836 859 C4-TYPE (POTENTIAL).
FT DOMAIN 2324 2343 C4-TYPE (POTENTIAL).
FT DOMAIN 4414 4493 POLY-PRO.
FT DOMAIN 4627 4726 PDZ.
FT DOMAIN 5003 5094 C2 DOMAIN 1.
FT DOMAIN 5120 AA; 560751 MW; A658D9891B65D412 CRC64;
SQ SEQUENCE

Query Match 7.7%; Score 578.5; DB 1; Length 5120;
Best Local Similarity 20.8%; Pred. No. 6.6e-15;
Matches 333; Conservative 179; Mismatches 542; Indels 549; Gaps 70;

Qy 108 HNPTSPSSKAPPPSGASQT---IKSTKSPKPKKTKKYISEETIEHSVSEN 163
Db 5 HHPRQP--GKPPDPGPGISKSRVTVLKEGRAP-----GRSPSISIRSEKSRDPEKD 57
Qy 164 QESSSSSSSSSSSSSTIMKIXS-----SKNSANRELOKKLYKONKKNRTK--- 209
Db 58 QKPSMMPSEFSEANPLAVTSVVKPSFPDLISDDAHERAGKQKTKQKEQGPEBQR 117
Qy 210 ---KKPTPK--PVLVDAGSGLDNGDFKVTTPDTSTTQHNKVSTSPKITTAKPINR--- 261
Db 118 GLAHNPGQGGPKVLQGG-----FVAPLTQGLE-SSKVPVQGGQPPSPKQGG 164
Qy 262 ---PSLPNSDTSKETSITVNKETVETKETTTNNKOTSDGKEKTTSAKETSISAKETS 318
Db 165 KPGSHHPGDSKAEQ-----VKQPPQPRGPKSQLOQSEPTPGQOQTSIA-----KTS 212
Qy 319 KDLAPTSVLAKPFPKATTTKGAALTTPKKEPTTPTEKPSSTTPKEPT----- 367
Db 213 ---GPT-----KLPDQPSAK-----TSSQAPPTPKSLQSGSVKQPSQPARQGPVK 260
Qy 368 PTTIKSAP--TTPKEPAPTTKSAPT--TPKEPAPTTT-----KEPAPTTKPEAPTTKEP 420
Db 261 PSAQAGAPKQKQPSSEKPTAQTGTAPAKQPPRGKPTPLQGTGVKQVPPAGPT---KP 317
Qy 421 APPTKSAPTPKKEPAPTP--KKPAPTTKEPAPTTKEPTTPKE---PAPTTKEPA 475
Db 318 SSQTAGAKSIAQOPGILTKPPGQOPGPEKPLQGRQASTTQGVESSTPKKTFCLCTTELL 377

```



```

QY 476 PTPK-----PAPAPK-----PAP 492
D 378 LHTPEKANTYTCQCHTVVSLGCFNPNHITELKELCLNCOMORALGSDLASGHCPCP 437
QY 493 TTPKEPAPTPKPEAPTTKEPSPT---TPKEPAPTTKSAPTTTKEPAPTTKSAPTT 548
D 438 QLE-----PKOKPTPTASTAKPSBOLQPGOKDASPRDPSQOADSCKVPF---QKQPSM 491
QY 549 PKEPSPPTTKEPAPTPPKKAPPTPKKAPPTK---EPAPTTKEPAPTTTKKAPTA 604
D 492 PGSP-EVKSQO-----THAEPSTP--GQOIDSTPKSDQVKTQAEKQNOISICKPTMDIV 544
QY 605 PKEPAPTTPEKETA-----PTPKLTPTTPEKLAPTTPEKAPPTPEELATTTBEPTPT 660
D 545 PTSAAGCVQODLADPOSPSTQOKVT---DSMPETTKPADTHPACDKDS---KLPPOV 558
QY 661 PEEBAPTPKAAAPNTPEKEBAPT-TPKEBAPT---TPK---EPAPTTPEKETAFTPK-GT 712
D 599 SRQ--KSDPKLASGSAKSDAKTOKPESEBAPVMDPKQTKAPAKPDTPKAPKQOAGT 656
QY 713 APPTLKEBAPTTKAPKAPKLAFTTTKEPTS-----TSDKAPPTPK-----755
D 657 GPR-----PTSAQAPAPPOQOKTPEGSRRLNCGITDAPKQPTTPOETVNGKLP 709
QY 756 -----GAPTTPEKAPPT-----PKEPAPTTKETA 782
D 710 GRCASIFSOASSLISTAGOGSOTSGAPAPATKOPPOPSOPASOAPKEAQAQPPPK 769
QY 783 PTTLKEBAPTPPKKAPKELAPT-TTKGP-----TSTSDKAPPTPEKETA 827
D 770 APFKKTKPLASEKLGPMASDSTLTITKGSDEKKPLAKDSKHQTLMAKKPALSQEK 829
QY 828 -----827
D 830 SOPKVSCLPCKTGLNIGSKDPNNTCTECKVAVNLCGFNPMHIVEOEWLCLNCQTO 889
QY 828 -----PTTKEBAPPTP-KKAPPTPEPTP-----PPT---TSEVST 859
D 890 RAMSGOLGDMGKVPKLGCBSPQSVSKRPATPOKQPVANSHSQSKSTPTPATPKPE 949
QY 860 PTTKEPPTTIHKSDESTBELSAEPTPKAL--ENSPEPSPVPTTKTTPAATKPEMTTAKD 917
D 950 PSVPKSVPLQOGKLEKT--LSADKIQOGIQEKDASKGS--KLFKTPSADKQVRSQKXD 1006
QY 918 -----KTERDLKTPETTTAAPKTKZ-----940
D 1007 SRLQOTKLTTPSSDKILNGVQEKDKFOEAKLAKIPSAADKILHRLQEKDPKLOQMKMAK 1066
QY 941 -----TATTEKTTES-----KITATTT-----Q 959
D 1067 ALSADKIQPEAQEDVQLOEVLRSKVASADKIOHGIOKDLNQHVKIEKTSVEKIQEAO 1126
QY 960 VTSTTQDTPPFKITTLKTTTLAPKTYTT-KKTIITTEIMNKE-----ETAKPKDRATNS 1014
D 1127 KSKKLOQDKLP--KTLSEDKIPATVSSHKDLLSSEBKKKLELEKSTPHKDKKEDQ 1183
QY 1015 KATT-----PKQKPTKAPKPTSTKQKPTMPVRVKKPTTPPRK 1055
D 1184 TALTGHTIEOKAVEAPCDKLEKQEDVKEDLTGTGIPOMVSKPEKKEEETKTPVP---1240
QY 1056 TSTMPELNPTSRIEAM-----LOTTTPNOTPNKILVENVKSDAGABEETHML 1109
D 1241 VSRLEPSDIVEAVREKIEKEDKSDTSSGQOQSPGLSDTGYSSGJISSJGEIP---1296
QY 1110 RPHVPEBVTYPPMDYLPVNVNOCIIINPMLSDETNINCNGPVNGLTTLNGLTLVAFRGHY 1169
D 1297 -----SHLPSDEKOLPREPSQ---KDLISQ-----1319
QY 1170 FMMLSFPSPSPARATTEWGISPIDYPTTRCNGCGKTPFFKDS-----QYRFPNDIK 1224
D 1320 -----SPSPSPS-----DLAKJESTYLSILEAQASTLTDEKSVKREKELYETSEQTK 1365

```

```

QY 1225 DAGYKPIFKGGLTQIUALSTAKYKNWPESVYFFK-----RGGSIOQYIK 1274
D 1366 DQHKTYEL-----EYTPESYSSDEEDLAIQEGERTTADSKGASSTQDYK 1412
QY 1275 QEB-VQKCPGR--PALNYPVYGEMTOYRRRRFERAIGSPOT 1313
D 1413 EEDGNDTPTARQRYDVSDESSSENSPVRRRRRASVSSSS 1455

RESULT 12
ZAN HUMAN
ID ZAN HUMAN STANDARD; PRT; 2812 AA.
AC Q9Y493; Q00218; Q96L85; Q96L86; Q96L87; Q96L88; Q96L89; Q96L90;
AC Q9BX99; Q9B283; Q9B284; Q9B285; Q9B286; Q9B287; Q9B288;
DT 16-OCT-2001 (Rel. 40, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Zonadhesin precursor.
GN ZAN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1; 2; 3; 4; 5 AND 6).
RC TISSUE=Testis;
RA Cheung T.L., Masiar M.J., Cornwall G.A., Hardy D.M.;
RT "Multiple intra-species variants of human zonadhesin."
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP PARTIAL SEQUENCE FROM N.A.
RA MEDLINE=99018118; PubMed=9799793;
RA Glockner G., Scherer S., Schatevov R., Boright A.P., Weber J.,
RA Tsui L.-C., Rosenthal A.;
RT "Large-scale sequencing of two regions in human chromosome 7q22:
RT analysis of 650 kb of genomic sequence around the EPO and CUGT1 loci
RT reveals 17 genes."
RL Genome Res. 8:1060-1073(1998).
RN [3]
RP SEQUENCE OF 1810-2812 FROM N.A. (ISOFORM 1).
RA MEDLINE=21138439; PubMed=11239002;
RA Wilson M.D., Riemer C., Martindale D.W., Schnupf P., Boright A.P.,
RA Cheung T.L., Hardy D.M., Schwartz S., Scherer S.W., Tsui L.-C.,
RA Miller W., Koop B.F.;
RT "Comparative analysis of the gene-dense ACHE/TFE2 region on human
RT chromosome 7q22 with the orthologous region on mouse chromosome 5."
RL Nucleic Acids Res. 29:1352-1365(2001).
RN [4]
RP SEQUENCE OF 2375-2683 FROM N.A. (ISOFORM 7).
RC TISSUE=Testis;
RA MEDLINE=97271566; PubMed=9126492;
RA Gao Z., Hartum T., Garbers D.L.;
RT "Chromosome localization of the mouse zonadhesin gene and the human
RT zonadhesin gene (ZAN)."
RL Genomics 41:119-122(1997).
CC - FUNCTION: BINDS IN A SPECIES-SPECIFIC MANNER TO THE ZONA PELLUCIDA
CC OF THE EGG. MAY BE INVOLVED IN GAMETE RECOGNITION AND/OR
CC SIGNALING.
CC - SUBUNIT: PROBABLY FORMS COVALENT OLIGOMERS.
CC - SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN, EXCLUSIVELY ON THE
CC APICAL REGION OF THE SPERM HEAD (BY SIMILARITY).
CC - ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=7;
CC Name=3;
CC IsoId=Q9Y493-1; Sequence=Displayed;
CC Name=1;
CC IsoId=Q9Y493-2; Sequence=VSP_001430, VSP_001431;
CC Name=2;
CC IsoId=Q9Y493-3; Sequence=VSP_001428, VSP_001429;
CC Name=4;
CC IsoId=Q9Y493-4; Sequence=VSP_001424, VSP_001425;
CC Name=5;
CC IsoId=Q9Y493-5; Sequence=VSP_001420, VSP_001421;

```

CC	Name=5;
CC	Isoid=q9Y493-6; Sequence=VSP_001422, VSP_001423;
CC	Name=7/;
CC	Isoid=q9Y493-7; Sequence=VSP_001426, VSP_001427;
CC	-1- TISSUE SPECIFICITY: IN TESTIS; PRIMARILY IN HAPLOID SPERMATIDS.
CC	-1- DOMAIN: THE MAM DOMAINS PROBABLY MEDIATE SPERM ADHESION TO THE
CC	ZONA PELLUCIDA.
CC	-1- DOMAIN: DURING SPERM MIGRATION THROUGH THE REPRODUCTIVE TRACTS,
CC	THE MUCIN-LIKE DOMAIN MIGHT INHIBIT INAPPROPRIATE TRAPPING OF
CC	SPERMATOZOA OR PROMOTING ADHESION TO THE OVIDUCTAL ISTHMUS.
CC	-1- DOMAIN: The VWFD domain 2 may mediate covalent oligomerization (by
CC	similarity to human intestinal mucin MUC2).
CC	-1- SIMILARITY: Contains 3 MAM domains.
CC	-1- SIMILARITY: Contains 4 VWFD domains.
CC	-1- SIMILARITY: Contains 1 EGF-like domain.
CC	-1- CAUTION: Ref.2 sequence differs from that shown due to
CC	transposition of a number of exons.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on the
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC	or send an email to license@isb-sib.ch).
CC	-----
DR	EMBL; AF332975; AAK01431.1; -
DR	EMBL; AF332976; AAK01432.1; -
DR	EMBL; AF332977; AAK01433.1; -
DR	EMBL; AF332978; AAK01434.1; -
DR	EMBL; AF332979; AAK01435.1; -
DR	EMBL; AF332980; AAK01436.1; -
DR	EMBL; AY046055; AAL04410.1; -
DR	EMBL; AY046055; AAL04411.1; -
DR	EMBL; AY046055; AAL04412.1; -
DR	EMBL; AY046055; AAL04413.1; -
DR	EMBL; AY046055; AAL04414.1; -
DR	EMBL; AY046055; AAL04415.1; -
DR	EMBL; AY053356; AAC78790.1; -
DR	EMBL; AF312033; AAK21011.1; -
DR	EMBL; U83191; AAC51208.1; -
DR	Genew: HGNC:12857; ZAN.
DR	MIM: 602372; -
DR	GO: GO:0016324; C:apical plasma membrane; NAS.
DR	GO: GO:0007333; P:binding of sperm to zona pellucida; NAS.
DR	GO: GO:0008037; P:cell recognition; NAS.
DR	GO: GO:0016337; P:cell-cell adhesion; NAS.
DR	InterPro: IPR006209; EGF like.
DR	InterPro: IPR000998; MAM_domain.
DR	InterPro: IPR002919; TIL_Cysrich.
DR	InterPro: IPR003328; TILA_Cysrich.
DR	InterPro: IPR001007; VWF_C.
DR	InterPro: IPR001846; VWF_D.
DR	Pfam: PF00629; MAM; 3.
DR	Pfam: PF01826; TIL; 4.
DR	Pfam: PF02345; TILA; 5.
DR	Pfam: PF00094; vwd; 4.
DR	SMART; SM00181; EGF; 4.
DR	SMART; SM00137; MAM; 3.
DR	SMART; SM00214; WMC; 4.
DR	SMART; SM00215; VWC_out; 4.
DR	SMART; SM00216; VMD; 4.
DR	PROSITE: PS00022; EGF_1; 1.
DR	PROSITE: PS01186; EGF_2; 4.
DR	PROSITE: PS00740; MAM_1; 1.
DR	PROSITE: PS50060; MAM_2; 3.
KW	Signal; Glycoprotein; Transmembrane; EGF-like domain; Cell adhesion;
KW	Repeat; Alternative splicing.
FT	SIGNAL 1 17
FT	CHAIN 18 2812
FT	DOMAIN 18 2757
FT	TRANSMEM 2758 2778
FT	DOMAIN 2779 2812
FT	POTENTIAL.
FT	ZONADHESIN.
FT	EXTRACELLULAR (POTENTIAL).
FT	POTENTIAL.
FT	CYTOPLASMIC (POTENTIAL).

FT	DOMAIN	39	204	MAM 1.
FT	DOMAIN	209	368	MAM 2.
FT	DOMAIN	371	536	MAM 3.
FT	DOMAIN	573	1041	66 X HEPTAPEPTIDE REPEATS (APPROXIMATE) (MUCIN-LIKE DOMAIN).
FT	DOMAIN	1156	1308	WVFED 1.
FT	DOMAIN	1542	1697	WVFED 2.
FT	DOMAIN	1931	2086	WVFED 3.
FT	DOMAIN	2331	2485	WVFED 4.
FT	DOMAIN	2708	2744	EGF-LIKE.
FT	DISULFID	2712	2723	BY SIMILARITY.
FT	DISULFID	2717	2732	BY SIMILARITY.
FT	DISULFID	2734	2743	BY SIMILARITY.
FT	CARBOHYD	333	333	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	493	493	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1112	1112	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1188	1188	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1685	1685	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1804	1804	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1900	1900	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1946	1946	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	2203	2203	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	2542	2542	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	2701	2701	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	VARSPLIC	2537	2601	HGVSS -> MAAAF (in isoform 5). /FTId=VSP 001420. Missing (in isoform 5). /FTId=VSP 001421.
FT	VARSPLIC	2602	2812	HGVSSRYHISELYDTLPISLIC -> YALICQAGALAGMR DRTC (in isoform 6). /FTId=VSP 001422. Missing (in isoform 6). /FTId=VSP 001422.
FT	VARSPLIC	2597	2617	HGVSSRYHISELYDTLPISLIC -> YALICQAGALAGMR DRTC (in isoform 6). /FTId=VSP 001423. Missing (in isoform 6). /FTId=VSP 001423.
FT	VARSPLIC	2618	2708	HGVSSRYHISELYDTLPISLICPGRRPG -> YALICQAG AALAGMRDRITCAGQQLSD (in isoform 4). /FTId=VSP 001424. Missing (in isoform 4). /FTId=VSP 001425.
FT	VARSPLIC	2597	2624	HGVSSRYHISELYDTLPISLICPGRRPG -> YALICQAG AALAGMRDRITCAGQQLSD (in isoform 4). /FTId=VSP 001424. Missing (in isoform 4). /FTId=VSP 001425.
FT	VARSPLIC	2625	2812	HGVSSRYHISELYDTLPISLICPGRRPGRLGRRLRQHP ASRLILCEPR -> YALICQAGALAGMRDRITCAMECPAGTIYOSCTPC PASCANLADPGCEPVCGCAD (in isoform 7). /FTId=VSP 001426. Missing (in isoform 7). /FTId=VSP 001427.
FT	VARSPLIC	2597	2636	HGVSSRYHISELYDTLPISLICPGRRPGRLGRRLRQHP ASRLILCEPR -> YALICQAGALAGMRDRITCAMECPAGTIYOSCTPC PASCANLADPGCEPVCGCAD (in isoform 7). /FTId=VSP 001426. Missing (in isoform 7). /FTId=VSP 001427.
FT	VARSPLIC	2663	2666	HGVSSRYHISELYDTLPISLICPGRRPGRLGRRLRQHP ASRLILCEPR -> YALICQAGALAGMRDRITCAMECPAGTIYOSCTPC PASCANLADPGCEPVCGCAD (in isoform 7). /FTId=VSP 001426. Missing (in isoform 7). /FTId=VSP 001427.
FT	VARSPLIC	2597	2689	HGVSSRYHISELYDTLPISLICPGRRPGRLGRRLRQHP ASRLILCEPR -> YALICQAGALAGMRDRITCAMECPAGTIYOSCTPC PASCANLADPGCEPVCGCAD (in isoform 7). /FTId=VSP 001426. Missing (in isoform 7). /FTId=VSP 001427.
FT	VARSPLIC	2597	2636	HGVSSRYHISELYDTLPISLICPGRRPGRLGRRLRQHP ASRLILCEPR -> YALICQAGALAGMRDRITCAMECPAGTIYOSCTPC PASCANLADPGCEPVCGCAD (in isoform 7). /FTId=VSP 001426. Missing (in isoform 7). /FTId=VSP 001427.
FT	VARSPLIC	2663	2666	HGVSSRYHISELYDTLPISLICPGRRPGRLGRRLRQHP ASRLILCEPR -> YALICQAGALAGMRDRITCAMECPAGTIYOSCTPC PASCANLADPGCEPVCGCAD (in isoform 7). /FTId=VSP 001426. Missing (in isoform 7). /FTId=VSP 001427.
FT	VARSPLIC	2597	2689	HGVSSRYHISELYDTLPISLICPGRRPGRLGRRLRQHP ASRLILCEPR -> YALICQAGALAGMRDRITCAMECPAGTIYOSCTPC PASCANLADPGCEPVCGCAD (in isoform 7). /FTId=VSP 001426. Missing (in isoform 7). /FTId=VSP 001427.
FT	VARSPLIC	2597	2636	HGVSSRYHISELYDTLPISLICPGRRPGRLGRRLRQHP ASRLILCEPR -> YALICQAGALAGMRDRITCAMECPAGTIYOSCTPC PASCANLADPGCEPVCGCAD (in isoform 7). /FTId=VSP 001426. Missing (in isoform 7). /FTId=VSP 001427.
FT	VARSPLIC	2663	2666	HGVSSRYHISELYDTLPISLICPGRRPGRLGRRLRQHP ASRLILCEPR -> YALICQAGALAGMRDRITCAMECPAGTIYOSCTPC PASCANLADPGCEPVCGCAD (in isoform 7). /FTId=VSP 001426. Missing (in isoform 7). /FTId=VSP 001427.
FT	VARSPLIC	2597	2689	HGVSSRYHISELYDTLPISLICPGRRPGRLGRRLRQHP ASRLILCEPR -> YALICQAGALAGMRDRITCAMECPAGTIYOSCTPC PASCANLADPGCEPVCGCAD (in isoform 7). /FTId=VSP 001426. Missing (in isoform 7). /FTId=VSP 001427.
FT	VARSPLIC	2597	2636	HGVSSRYHISELYDTLPISLICPGRRPGRLGRRLRQHP ASRLILCEPR -> YALICQAGALAGMRDRITCAMECPAGTIYOSCTPC PASCANLADPGCEPVCGCAD (in isoform 7). /FTId=VSP 001426. Missing (in isoform 7). /FTId=VSP 001427.
FT	VARSPLIC	2663	2666	HGVSSRYHISELYDTLPISLICPGRRPGRLGRRLRQHP ASRLILCEPR -> YALICQAGALAGMRDRITCAMECPAGTIYOSCTPC PASCANLADPGCEPVCGCAD (in isoform 7). /FTId=VSP 001426. Missing (in isoform 7). /FTId=VSP 001427.
FT	VARSPLIC	2597	2689	HGVSSRYHISELYDTLPISLICPGRRPGRLGRRLRQHP ASRLILCEPR -> YALICQAGALAGMRDRITCAMECPAGTIYOSCTPC PASCANLADPGCEPVCGCAD (in isoform 7). /FTId=VSP 001426. Missing (in isoform 7). /FTId=VSP 001427.
FT	VARSPLIC	2597	2636	HGVSSRYHISELYDTLPISLICPGRRPGRLGRRLRQHP ASRLILCEPR -> YALICQAGALAGMRDRITCAMECPAGTIYOSCTPC PASCANLADPGCEPVCGCAD (in isoform 7). /FTId=VSP 001426. Missing (in isoform 7). /FTId=VSP 001427.
FT	VARSPLIC	2663	2666	HGVSSRYHISELYDTLPISLICPGRRPGRLGRRLRQHP ASRLILCEPR -> YALICQAGALAGMRDRITCAMECPAGTIYOSCTPC PASCANLADPGCEPVCGCAD (in isoform 7). /FTId=VSP 001426. Missing (in isoform 7). /FTId=VSP 001427.
FT	VARSPLIC	2597	2689	HGVSSRYHISELYDTLPISLICPGRRPGRLGRRLRQHP ASRLILCEPR -> YALICQAGALAGMRDRITCAMECPAGTIYOSCTPC PASCANLADPGCEPVCGCAD (in isoform 7). /FTId=VSP 001426. Missing (in isoform 7). /FTId=VSP 001427.
FT	VARSPLIC	2597	2636	HGVSSRYHISELYDTLPISLICPGRRPGRLGRRLRQHP ASRLILCEPR -> YALICQAGALAGMRDRITCAMECPAGTIYOSCTPC PASCANLADPGCEPVCGCAD (in isoform 7). /FTId=VSP 001426. Missing (in isoform 7). /FTId=VSP 001427.
FT	VARSPLIC	2663	2666	HGVSSRYHISELYDTLPISLICPGRRPGRLGRRLRQHP ASRLILCEPR -> YALICQAGALAGMRDRITCAMECPAGTIYOSCTPC PASCANLADPGCEPVCGCAD (in isoform 7). /FTId=VSP 001426. Missing (in isoform 7). /FTId=VSP 001427.
FT	VARSPLIC	2597	2689	HGVSSRYHISELYDTLPISLICPGRRPGRLGRRLRQHP ASRLILCEPR -> YALICQAGALAGMRDRITCAMECPAGTIYOSCTPC PASCANLADPGCEPVCGCAD (in isoform 7). /FTId=VSP 001426. Missing (in isoform 7). /FTId=VSP 001427.
FT	VARSPLIC	2597	2636	HGVSSRYHISELYDTLPISLICPGRRPGRLGRRLRQHP ASRLILCEPR -> YALICQAGALAGMRDRITCAMECPAGTIYOSCTPC PASCANLADPGCEPVCGCAD (in isoform 7). /FTId=VSP 001426. Missing (in isoform 7). /FTId=VSP 001427.
FT	VARSPLIC	2663	2666	HGVSSRYHISELYDTLPISLICPGRRPGRLGRRLRQHP ASRLILCEPR -> YALICQAGALAGMRDRITCAMECPAGTIYOSCTPC PASCANLADPGCEPVCGCAD (in isoform 7). /FTId=VSP 001426. Missing (in isoform 7). /FTId=VSP 001427.
FT	VARSPLIC	2597	2689	HGVSSRYHISELYDTLPISLICPGRRPGRLGRRLRQHP ASRLILCEPR -> YALICQAGALAGMRDRITCAMECPAGTIYOSCTPC PASCANLADPGCEPVCGCAD (in isoform 7). /FTId=VSP 001426. Missing (in isoform 7). /FTId=VSP 001427.
FT	VARSPLIC	2597	2636	HGVSSRYHISELYDTLPISLICPGRRPGRLGRRLRQHP ASRLILCEPR -> YALICQAGALAGMRDRITCAMECPAGTIYOSCTPC PASCANLADPGCEPVCGCAD (in isoform 7). /FTId=VSP 001426. Missing (in isoform 7). /FTId=VSP 001427.
FT	VARSPLIC	2663	2666	HGVSSRYHISELYDTLPISLICPGRRPGRLGRRLRQHP ASRLILCEPR -> YALICQAGALAGMRDRITCAMECPAGTIYOSCTPC PASCANLADPGCEPVCGCAD (in isoform 7). /FTId=VSP 001426. Missing (in isoform 7). /FTId=VSP 001427.
FT	VARSPLIC	2597	2689	HGVSSRYHISELYDTLPISLICPGRRPGRLGRRLRQHP ASRLILCEPR -> YALICQAGALAGMRDRITCAMECPAGTIYOSCTPC PASCANLADPGCEPVCGCAD (in isoform 7). /FTId=VSP 001426. Missing (in isoform 7). /FTId=VSP 001427.
FT	VARSPLIC	2597	2636	HGVSSRYHISELYDTLPISLICPGRRPGRLGRRLRQHP ASRLILCEPR -> YALICQAGALAGMRDRITCAMECPAGTIYOSCTPC PASCANLADPGCEPVCGCAD (in isoform 7). /FTId=VSP 001426. Missing (in isoform 7). /FTId=VSP 001427.
FT	VARSPLIC	2663	2666	HGVSSRYHISELYDTLPISLICPGRRPGRLGRRLRQHP ASRLILCEPR -> YALICQAGALAGMRDRITCAMECPAGTIYOSCTPC PASCANLADPGCEPVCGCAD (in isoform 7). /FTId=VSP 001426. Missing (in isoform 7). /FTId=VSP 001427.
FT	VARSPLIC	2597	2689	HGVSSRYHISELYDTLPISLICPGRRPGRLGRRLRQHP ASRLILCEPR -> YALICQAGALAGMRDRITCAMECPAGTIYOSCTPC PASCANLADPGCEPVCGCAD (in isoform 7). /FTId=VSP 001426. Missing (in isoform 7). /FTId=VSP 001427.
FT	VARSPLIC	2597	2636	HGVSSRYHISELYDTLPISLICPGRRPGRLGRRLRQHP ASRLILCEPR -> YALICQAGALAGMRDRITCAMECPAGTIYOSCTPC PASCANLADPGCEPVCGCAD (in isoform 7). /FTId=VSP 001426. Missing (in isoform 7). /FTId=VSP 001427.
FT	VARSPLIC	2663	2666	HGVSSRYHISELYDTLPISLICPGRRPGRLGRRLRQHP ASRLILCEPR -> YALICQAGALAGMRDRITCAMECPAGTIYOSCTPC PASCANLADPGCEPVCGCAD (in isoform 7). /FTId=VSP 001426. Missing (in isoform 7). /FTId=VSP 001427.
FT	VARSPLIC	2597	2689	HGVSSRYHISELYDTLPISLICPGRRPGRLGRRLRQHP ASRLILCEPR -> YALICQAGALAGMRDRITCAMECPAGTIYOSCTPC PASCANLADPGCEPVCGCAD (in isoform 7). /FTId=VSP 001426. Missing (in isoform 7). /FTId=VSP 001427.
FT	VARSPLIC	2597	2636	HGVSSRYHISELYDTLPISLICPGRRPGRLGRRLRQHP ASRLILCEPR -> YALICQAGALAGMRDRITCAMECPAGTIYOSCTPC PASCANLADPGCEPVCGCAD (in isoform 7). /FTId=VSP 001426. Missing (in isoform 7). /FTId=VSP 001427.
FT	VARSPLIC	2663	2666	HGVSSRYHISELYDTLPISLICPGRRPGRLGRRLRQHP ASRLILCEPR -> YALICQAGALAGMRDRITCAMECPAGTIYOSCTPC PASCANLADPGCEPVCGCAD (in isoform 7). /FTId=VSP 001426. Missing (in isoform 7). /FTId=VSP 001427.
FT	VARSPLIC	2597	2689	HGVSSRYHISELYDTLPISLICPGRRPGRLGRRLRQHP ASRLILCEPR -> YALICQAGALAGMRDRITCAMECPAGTIYOSCTPC PASCANLADPGCEPVCGCAD (in isoform 7). /FTId=VSP 001426. Missing (in isoform 7). /FTId=VSP 001427.
FT	VARSPLIC	2597	2636	HGVSSRYHISELYDTLPISLICPGRRPGRLGRRLRQHP ASRLILCEPR -> YALICQAGALAGMRDRITCAMECPAGTIYOSCTPC PASCANLADPGCEPVCGCAD (in isoform 7). /FTId=VSP 001426. Missing (in isoform 7). /FTId=VSP 001427.
FT	VARSPLIC	2663	2666	HGVSSRYHISELYDTLPISLICPGRRPGRLGRRLRQHP ASRLILCEPR -> YALICQAGALAGMRDRITCAMECPAGTIYOSCTPC PASCANLADPGCEPVCGCAD (in isoform 7). /FTId=VSP 001426. Missing (in isoform 7). /FTId=VSP 001427.
FT	VARSPLIC	2597	2689	HGVSSRYHISELYDTLPISLICPGRRPGRLGRRLRQHP ASRLILCEPR -> YALICQAGALAGMRDRITCAMECPAGTIYOSCTPC PASCANLADPGCEPVCGCAD (in isoform 7). /FTId=VSP 001426. Missing (in isoform 7). /FTId=VSP 001427.
FT	VARSPLIC	2597	2636	HGVSSRYHISELYDTLPISLICPGRRPGRLGRRLRQHP ASRLILCEPR -> YALICQAGALAGMRDRITCAMECPAGTIYOSCTPC PASCANLADPGCEPVCGCAD (in isoform 7). /FTId=VSP 001426. Missing (in isoform 7). /FTId=VSP 001427.
FT	VARSPLIC	2663	2666	HGVSSRYHISELYDTLPISLICPGRRPGRLGRRLRQHP ASRLILCEPR -> YALICQAGALAGMRDRITCAMECPAGTIYOSCTPC PASCANLADPGCEPVCGCAD (in isoform 7). /FTId=VSP 001426. Missing (in isoform 7). /FTId=VSP 001427.
FT	VARSPLIC	2597	2689	HGVSSRYHISELYDTLPISLICPGRRPGRLGRRLRQHP ASRLILCEPR -> YALICQAGALAGMRDRITCAMECPAGTIYOSCTPC PASCANLADPGCEPVCGCAD (in isoform 7). /FTId=VSP 001426. Missing (in isoform 7). /FTId=VSP 001427.
FT	VARSPLIC	2597	2636	HGVSSRYHISELYDTLPISLICPGRRPGRLGRRLRQHP ASRLILCEPR -> YALICQAGALAGMRDRITCAMECPAGTIYOSCTPC PASCANLADPGCEPVCGCAD (in isoform 7). /FTId=VSP 001426. Missing (in isoform 7). /FTId=VSP 001427.
FT	VARSPLIC	2663	2666	HGVSSRYHISELYDTLPISLICPGRRPGRLGRRLRQHP ASRLILCEPR -> YALICQAGALAGMRDRITCAMECPAGTIYOSCTPC PASCANLADPGCEPVCGCAD (in isoform 7). /FTId=VSP 001426. Missing (in isoform 7). /FTId=VSP 001427.
FT	VARSPLIC	2597	2689	HGVSSRYHISELYDTLPISLICPGRRPGRLGRRLRQHP ASRLILCEPR -> YALICQAGALAGMRDRITCAMECPAGTIYOSCTPC PASCANLADPGCEPVCGCAD (in isoform 7). /FTId=VSP 001426. Missing (in isoform 7). /FTId=VSP 001427.
FT	VARSPLIC	2597	2636	HGVSSRYHISELYDTLPISLICPGRRPGRLGRRLRQHP ASRLILCEPR -> YALICQAGALAGMRDRITCAMECPAGTIYOSCTPC PASCANLADPGCEPVCGCAD (in isoform 7). /FTId=VSP 001426. Missing (in isoform 7). /FTId=VSP 001427.
FT	VARSPLIC	2663	2666	HGVSSRYHISELYDTLPISLICPGRRPGRLGRRLRQHP ASRLILCEPR -> YALICQAGALAGMRDRITCAMECPAGTIYOSCTPC PASCANLADPGCEPVCGCAD (in isoform 7). /FTId=VSP 001426. Missing (in isoform 7). /FTId=VSP 001427.
FT	VARSPLIC	2597	2689	HGVSSRYHISELYDTLPISLICPGRRPGRLGRRLRQHP ASRLILCEPR -> YALICQAGALAGMRDRITCAMECPAGTIYOSCTPC PASCANLADPGCEPVCGCAD (in isoform 7). /FTId=VSP 001426. Missing (in isoform 7). /FTId=VSP 001427.
FT	VARSPLIC	2597	2636	HGVSSRYHISELYDTLPISLICPGRRPGRLGRRLRQHP ASRLILCEPR -> YALICQAGALAGMRDRITCAMECPAGTIYOSCTPC PASCANLADPGCEPVCGCAD (in isoform 7). /FTId=VSP 001426. Missing (in isoform 7). /FTId=VSP 001427.
FT	VARSPLIC	2663	2666	HGVSSRYHISELYDTLPISLICPGRRPGRLGRRLRQHP ASRLILCEPR -> YALICQAGALAGMRDRITCAMECPAGTIYOSCTPC PASCANLADPGCEPVCGCAD (in isoform 7). /FTId=VSP 001426. Missing (in isoform 7). /FTId=VSP 001427.
FT	VARSPLIC	2597	2689	HGVSSRYHISELYDTLPISLICPGRRPGRLGRRLRQHP ASRLILCEPR -> YALICQAGALAGMRDRITCAMECPAGTIYOSCTPC PASCANLADPGCEPVCGCAD (in isoform 7). /FTId=VSP 001426. Missing (in isoform 7). /FTId=VSP 001427.
FT	VARSPLIC	2597	2636	HGVSSRYHISELYDTLPISLICPGRRPGRLGRRLRQHP ASRLILCEPR -> YALICQAGALAGMRDRITCAMECPAGTIYOSCTPC PASCANLADPGCEPVCGCAD (in isoform 7). /FTId=VSP 001426. Missing (in isoform 7). /FTId=VSP 001427.
FT	VARSPLIC	2663	2666	HGVSSRYHISELYDTLPISLICPGRRPGRLGRRLRQHP ASRLILCEPR -> YALICQAGALAGMRDRITCAMECPAGTIYOSCTPC PASCANLADPGCEPVCGCAD (in isoform 7). /FTId=VSP 001426. Missing (in isoform 7). /FTId=VSP 001427.
FT	VARSPLIC	2597	2689	HGVSSRYHISELYDTLPISLICPGRRPGRLGRRLRQHP ASRLILCEPR -> YALICQAGALAGMRDRITCAMECPAGTIYOSCTPC PASCANLADPGCEPVCGCAD (in isoform 7). /FTId=VSP 001426. Missing (in isoform 7). /FTId=VSP 001427.
FT	VARSPLIC	2597	2636	HGVSSRYHISELYDTLPISLICPGRRPGRLGRRLRQHP ASRLILCEPR -> YALICQAGALAGMRDRITCAMECPAGTIYOSCTPC PASCANLADPGCEPVCGCAD (in isoform 7). /FTId=VSP 001426. Missing (in isoform 7). /FTId=VSP 001427.
FT	VARSPLIC	2663	2666	HGVSSRYHISELYDTLPISLICPGRRPGRLGRRLRQHP ASRLILCEPR -> YALICQAGALAGMRDRITCAMECPAGTIYOSCTPC PASCANLADPGCEPVCGCAD (in isoform 7). /FTId=VSP 001426. Missing (in isoform 7). /FTId=VSP 001427.
FT	VARSPLIC	2597	2689	HGVSSRYHISELYDTLPISLICPGRRPGRLGRRLRQHP ASRLILCEPR -> YALICQAGALAGMRDRITCAMECPAGTIYOSCTPC PASCANLADPGCEPVCGCAD (in isoform 7). /FTId=VSP 001426. Missing (in isoform 7). /FTId=VSP 001427.
FT	VARSPLIC	2597	2636	HGVSSRYHISELYDTLPISLICPGRRPGRLGRRLRQHP ASRLILCEPR -> YALICQAGALAGMRDRITCAMECPAGTIYOSCTPC PASCANLADPGCEPVCGCAD (in isoform 7). /FTId=VSP 001426. Missing (in isoform 7). /FTId=VSP 001427.
FT	VARSPLIC	2663	2666	HGVSSRYHISELYDTLPISLICPGRRPGRLGRRLRQHP ASRLILCEPR -> YALICQAGALAGMRDRITCAMECPAGTIYOSCTPC PASCANLADPGCEPVCGCAD (in isoform 7). /FTId=VSP 001426. Missing (in isoform 7). /FTId=VSP 001427.
FT	VARSPLIC	2597	2689	HGVSSRYHISELYDTLPISLICPGRRPGRLGRRLRQHP ASRLILCEPR -> YALICQAGALAGMRDRITCAMECPAGTIYOSCTPC PASCANLADPGCEPVCGCAD (in isoform 7). /FTId=VSP 001426. Missing (in isoform 7). /FTId=VSP 001427.
FT	VARSPLIC	2597	2636	HGVSSRYHISELYDTLPISLICPGRRPGRLGRRLRQHP ASRLILCEPR -> YALICQAGALAGMRDRITCAMECPAGTIYOSCTPC PASCANLADPGCEPVCGCAD (in isoform 7). /FTId=VSP 001426. Missing (in isoform 7). /FTId=VSP 001427.
FT	VARSPLIC	2663	2666	HGVSSRYHISELYDTLPISLICPGRRPGRLGRRLRQHP ASRLILCEPR -> YALICQAGALAGMRDRITCAMECPAGTIYOSCTPC PASCANLADPGCEPVCGCAD (in isoform 7). /FTId=VSP 001426. Missing (in isoform 7). /FTId=VSP 001427.
FT	VARSPLIC	2597	2689	HGVSSRYHISELYDTLPISLICPGRRPGRLGRRLRQHP ASRLILCEPR -> YALICQAGALAGMRDRITCAMECPAGTIYOSCTPC PASCANLADPGCEPVCGCAD (in isoform 7). /FTId=VSP 001426. Missing (in isoform 7). /FTId=VSP 001427.
FT	VARSPLIC	2597	2636	HGVSSRYHISELYDTLPISLICPGRRPGRLGRRLRQHP ASRLILCEPR -> YALICQAGALAGMRDRITCAMECPAGTIYOSCTPC PASCANLADPGCEPVCGCAD (in isoform 7). /FTId=VSP 001426. Missing (in isoform 7). /FTId=VSP 001427.
FT	VARSPLIC	2663	2666	HGVSSRYHISELYDTLPISLICPGRRPGRLGRRLRQHP ASRLILCEPR -> YALICQAGALAGMRDRITCAMECPAGTIYOSCTPC PASCANLADPGCEPVCGCAD (in isoform 7). /FTId=VSP 001426. Missing (in isoform 7). /FTId=VSP 001427.
FT	VARSPLIC	2597	2689	HGVSSRYHISELYDTLPISLICPGRRPGRLGRRLRQHP ASRLILCEPR -> YALICQAGALAGMRDRITCAMECPAGTIYOSCTPC PASCANLADPGCEPVCGCAD (in isoform 7). /FTId=VSP 001426. Missing (in isoform 7). /FTId=VSP 001427.
FT	VARSPLIC	2597	2636	HGVSSRYHISELYDTLPISLICPGRRPGRLGRRLRQHP ASRLILCEPR -> YALICQAGALAGMRDRITCAMECPAGTIYOSCTPC PASCANLADPGCEPVCGCAD (in isoform 7). /FTId=VSP 001426. Missing (in isoform 7). /FTId=VSP 001427.
FT	VARSPLIC	2663	2666	HGVSSRYHISELYDTLPISLICPGRRPGRLGRRLRQHP ASRLILCEPR -> YALICQAGALAGMRDRITCAMECPAGTIYOSCTPC PASCANLADPGCEPVCGCAD (in isoform 7). /FTId=VSP 001426. Missing (in isoform 7). /FTId=VSP 001427.
FT	VARSPLIC	2597	2689	HGVSSRYHISELYDTLPISLICPGRRPGRLGRRLRQHP ASRLILCEPR -> YALICQAGALAGMRDRITCAMECPAGTIYOSCTPC PASCANLADPGCEPVCGCAD (in isoform 7). /FTId=VSP 001426. Missing (in isoform 7). /FTId=VSP 001427.
FT	VARSPLIC	2597	2636	HGVSSRYHISELYDTLPISLICPGRRPGRLGRRLRQHP ASRLILCEPR -> YALICQAGALAGMRDRITCAMECPAGTIYOSCTPC PASCANLADPGCEPVCGCAD (in isoform 7). /FTId=VSP 001426. Missing (in isoform 7). /FTId=VSP 001427.
FT	VARSPLIC	2663	2666	HGVSSRYHISELYDTLPISLICPGRRPGRLGRRLRQHP ASRLILCEPR -> YALICQAGALAGMRDRITCAMECPAGTIYOSCTPC PASCANLADPGCEPVCGCAD (in isoform 7). /FTId=VSP 001426. Missing (in isoform 7). /FTId=VSP 001427.
FT	VARSPLIC	2597	2689	HGVSSRYHISELYDTLPISLICPGRRPGRLGRRLRQHP ASRLILCEPR -> YALICQAGALAGMRDRITCAMECPAGTIYOSCTPC PASCANLADPGCEPVCGCAD (in isoform 7). /FTId=VSP 001426. Missing (in isoform 7). /FTId=VSP 001427.
FT	VARSPLIC	2597	2636	HGVSSRYHISELYDTLPISLICPGRRPGRLGRRLRQHP ASRLILCEPR ->

```

QY 317 SAKDLAPTSKVLAKPTPKAETTTGPAALTTPKEPTTPPKAPASTTPKEPT-----PTTI 371
DB 578 SVTTEKPT-----VPEKEPTIPTEKPTISTEK---PTIPEKKNMSEKTIPTSEKPTII 629
QY 372 KSAPPTPKKAPATTTKSAPPTPKKAPATTTKKEBAPTTPKKAAPTTPKKAAPTTPKKAAPT 430
DB 630 TEKPTIPSE-KPTIPSEKPTISTEKPTVPTPE--PTTPEETTTSMEEPTIPTEKSIPT 666
QY 431 TPKEBAPTTPKKAAPTTPKKAAPTTPKEPTTPPKKAPATTTKKAAPTTPKKAAPTTPKKAAPT 490
DB 687 --EKPSIPTEK---PTISMEETIISTEKPTISPEKPTIPEK---PTIPEKSTISPEK- 737
QY 491 APTTPEBAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPT 550
DB 738 -PTTPE-KPTIPE-KPTIPE-KPTIPE-KPTIPE-KPTIPE-KPTIPE-KPTIPE-KPTIPE-KPTIPE 791
QY 551 EBPPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTP 605
DB 792 EKPTISTEE--PTTPEETIISTEKPSIPEKPTIPEETTTSTETKLTIPM-- 847
QY 606 KEBAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTP 655
DB 848 EKPTISTEKPTIPEKPTISTEKPTIPEKPTIPEKPTIPEKPTIPEKPTIPEKPTIPEKPTIPEK 904
QY 656 PTTTPEBAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPT 715
DB 905 KPTIPEKPTISTEK-----PTTPE-KPTIPE-----ETIISTEKLTIP 945
QY 716 TKEBAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTP 769
DB 946 --EKPTISTEKPTIPEKPTISTEKPTIPEKPTIPEKPTIPEKPTIPEKPTIPEKPTIPEKPTIPE 1001
QY 770 PKKAPATTTKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTP 829
DB 1002 PHPSPTA-TGLALVMSPHAPSTPMTSV---ILGTTTSSSTSTECPPNARYESCACPA 1057
QY 830 TPKEBAPT 837
DB 1058 SCKSPAP 1065

RESULT 13
PPI_MYTBD
ID_PPI_MYTBD STANDARD; PRT; 875 AA.
AC 023460;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Adhesive plaque matrix protein (Polypheholic adhesive protein) (Foot
protein 1) (MERP1) (Fragment).
GN PPI.
OS Mytilus edulis (Blue mussel).
OC Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Mytiloida;
OC Mytiloidea; Mytilidae; Mytilus.
OX NCBI_TaxId=6550;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91025829; PubMed=1367451;
RA Filippa D.R., Lee S.M., Link R.P., Straubeberg S.L., Straubeberg R.L.;
RT "Structural and functional repetition in a marine mussel adhesive
protein.";
RL Biotechnol. Prog. 6:171-177(1990).
RN [2]
RP PARTIAL SEQUENCE, AND POST-TRANSLATIONAL MODIFICATIONS.
RX MEDLINE=83135732; PubMed=6298211;
RA Waite J.H.;
RT "Evidence for a repeating 3,4-dihydroxyphenylalanine- and
hydroxyproline-containing decapeptide in the adhesive protein of the
mussel, Mytilus edulis L.";
RL J. Biol. Chem. 258:2911-2915(1983).
CC -1- FUNCTION: PROVIDES ADHESIVENESS TO THE MUSSEL'S FOOT. MUSSELS
PRODUCE ONE OF THE STRONGEST WATER INSOLUBLE GLUE. THE MUSSEL'S

```

```

CC ADHESIVE IS A BUNDLE OF THREADS, CALLED A BYSSUS, FORMED BY A
CC FIBROUS COLLAGENOUS CORE COATED WITH ADHESIVE PROTEINS.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: PRODUCED BY THE BYSSAL GLAND.
CC -1- DOMAIN: ALMOST EXCLUSIVELY COMPOSED OF REPEATS OF A DECAPEPTIDE.
CC -1- PTM: THE DECAPEPTIDE A-K-P-S-Y-P-T-Y-K IS POST-TRANSLATIONALLY
CC MODIFIED AS FOLLOWING: THE SIXTH AND SEVENTH RESIDUES ARE
CC HYDROXYLATED AND THE PENULTIMATE IS A 3,4-DIHYDROXYPHENYLALANINE
CC (DOPA) DERIVED FROM TYROSINE.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation.
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb.slb.ch).
CC -----
CC DR EMBL; X54422; CAA38294.1; -.
CC DR PIR; S23760; S23760.
CC DR InterPro; IPR002964; Adhesive_dlaq.
CC DR InterPro; IPR006031; XYPEX.
CC DR Pfam; PF02162; XYPEX; 55.
CC DR PRINTS; PR01216; ADHESIVEI.
CC KM Repeat; Hydroxylation.
CC FT NON TER 1
CC FT DOMAIN 67 870
CC FT TANDEM REPEATS OF Y-K-[PA]-K-[LP]-[ST]-Y-
CC P-P-[ST].
CC SQ SEQUENCE 875 AA; 100412 MM; 6E8A5312748CACE CRC64;
CC -----
CC Query Match 7.4%; Score 559; DB 1; Length 875;
CC Best local Similarity 28.7%; Pred. NO. 8.4e-15;
CC Matches 272; Conservative 106; Mismatches 373; Indels 195; Gaps 56;
CC -----
QY 241 TQHNKVSISPKITAKPINP-----RP-SLPNSDTSKETSLTVNKEETTVET 286
DB 1 TKEBAPYKVKTSYSAKPYKPTTQPLKKKVDRPFTKSYPTTG-SKTNVLLPLAKKLSYK 59
QY 287 KETTTNKOTSDG--KEKTT--SAKETQSIEKTSANDLAFTSKVLAKPTPKAETTTG 341
DB 60 PIKTTYNAKNTNPPYKPKMTYPTPKPSYPTPKSKPTKYKPTIPTYKAKPSY- 117
QY 342 PALTPKEPTTPPE-----PASTTPKEPTPTTISAPTP-----KEBAPTTPSAAPT 392
DB 118 PSSYKPKTYPTTYPKLTYPPTYKPSYPTTYKPSYPTTYKPSYPTTYKPSYPTTYKPSY 177
QY 393 PKKAPATTTKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTP 439
DB 178 P-----PTYKAKPSYPTTYKAKPSYPTTYKAKPTTYKAKPTTYKAKPSYPTTYKAKPSY 233
QY 440 PKKPA--PTTKEBPA--PTTKEBPT--PTTKEB--APTKEBAPTTPKKAAPTTPKKAAPTTP 488
DB 234 KAKPSYPTTYKAKPSYPTTYKAKPSYPTTYKAKPSYPTTYKAKPSYPTTYKAKPSYPTTYKAK 293
QY 489 KPA--PTTKEBPA--PTTKEBPA--PTTKEBS--PTTKEBPA--PTTSAAPTTPKKAAPTTP 538
DB 294 KPSYPTTYKAKPSYPTTYKAKPSYPTTYKAKPSYPTTYKAKPSYPTTYKAKPSYPTTYKAKPSY 353
QY 539 PTTTSAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTP 594
DB 354 PTTYKAKPSYPTTYKAKPSYPTTYKAKPSYPTTYKAKPSYPTTYKAKPSYPTTYKAKPSYPTTYKAK 409
QY 595 TTTTKEBAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTP 654
DB 410 PPTYKAKPSYKAP-----TYPST-YKAKPSYPTTYKAKPSYPTTYKAKPSYPTTYKAKPSY 457
QY 655 EPTTPEBAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTP 708
DB 458 KPT-----YSTTKA-----KPSYKAKPSYPTTYKAKPSYPTTYKAKPSYPTTYKAKPSYPTTYKAK 506
QY 709 PKGTAPTTLKEBAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTP 764
DB 507 PLTYKPTTYK-PKPSYPTTYKPTTYPTTYK-----PKSYPTTYKAKPSY 551

```


RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89121513; PubMed=3220257;
 RA Julien J.-P., Cote F., Beaudet L., Sidky M., Flavell D., Grosved F.,
 RA Moshynski W.,
 RT "Sequence and structure of the mouse gene coding for the largest
 RT neurofilament subunit.";
 RL Gene 68:307-314 (1988).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89089138; PubMed=3145094;
 RA Shneidman P.S., Carden M.J., Lees J.F., Lazzarini R.A.;
 RT "The structure of the largest murine neurofilament protein (NF-H) as
 RT revealed by cDNA and genomic sequences.";
 RL Brain Res. 464:217-223 (1988).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Swiss Webster; TISSUE=Brain;
 RA Carden M.J.;
 RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: I, M,
 CC AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.
 CC NF-H HAS AN IMPORTANT FUNCTION IN MATURE AXONS THAT IS NOT
 CC SUBSERVED BY THE TWO SMALLER NF PROTEINS.
 CC -1- PTM: THERE ARE A NUMBER OF REPEATS OF THE TRIPEPTIDE K-S-P, NFH IS
 CC PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS
 CC THOUGHT THAT PHOSPHORYLATION OF NFH RESULTS IN THE FORMATION OF
 CC INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE
 CC OF AXONAL CALIBER.
 CC -1- PTM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING
 CC OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H), THE
 CC LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND
 CC COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.
 CC -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
 CC -1- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN IN POSITIONS 534
 CC TO 716 AND IS SHORTER DUE TO FRAMESHIFTS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL, M24496; AAA39813.1; -;
 DR EMBL, M23149; AAA39813.1; JOINED.
 DR EMBL, M24494; AAA39813.1; JOINED.
 DR EMBL, M24495; AAA39813.1; JOINED.
 DR EMBL, M35131; AAA39809.1; ALT_FRAME.
 DR EMBL, Z31012; CAA83222.1; -;
 DR PIR, J03068; QPMH.
 DR MGD, MGI:97309; Neft.
 DR InterPro; IPR001664; IF.
 DR Pfam; PF00038; filament; 1.
 DR PROSITE; PS00226; IF; 1.
 KM Intermediate filament; Coiled coil; Neurone; Phosphorylation;
 Repeat.
 FT 1 97 HEAD.
 FT DOMAIN 98 408 ROD.
 FT DOMAIN 409 1087 TAIL.
 FT DOMAIN 436 517 GLU-RICH (ACIDIC).
 FT DOMAIN 519 865 50 X 6 AA TANDDEM REPEATS OF K-S-P-A-E-A.
 FT DOMAIN 887 1087 GLU/LYS-RICH.
 FT DOMAIN 98 129 COIL 1A.
 FT DOMAIN 130 141 COIL 1B.
 FT DOMAIN 142 239 COIL 1B.
 FT DOMAIN 240 261 LINKER 12.
 FT DOMAIN 262 283 COIL 2A.
 FT DOMAIN 284 287 LINKER 2.
 FT DOMAIN 288 408 COIL 2B.
 FT DOMAIN 133 133 K -> QA (IN REF. 2 AND 3).
 FT CONFLICT 133 133 A -> AR (IN REF. 2 AND 3).
 FT CONFLICT 199 199

FT CONFLICT 281 281 S -> T (IN REF. 2 AND 3).
 FT CONFLICT 492 492 L -> G (IN REF. 2 AND 3).
 FT CONFLICT 551 551 P -> PREAKS (IN REF. 3).
 FT CONFLICT 689 712 MISSING (IN REF. 3).
 FT CONFLICT 714 714 G -> A (IN REF. 3).
 FT CONFLICT 814 814 V -> M (IN REF. 2 AND 3).
 FT CONFLICT 843 843 T -> N (IN REF. 2 AND 3).
 SQ SEQUENCE 1087 AA; 116612 MW; 57BAC76A38ED1CB9 CRC64;
 Query Match 7.3%; Score 551; DB 1; Length 1087;
 Best Local Similarity 26.8%; Pred. No. 2e-14;
 Matches 224; Conservative 96; Mismatches 358; Indels 186; Gaps 43;
 143 KTKKVIIESE--EITEHSVS--ENESSSSSSSSSSSTIWKIKSKNSANRELQK 199
 331 KSTKESLEKQSELEDRHOADIASYODAIQOLDSLEIRTKWM-----AAQREYODLIN 385
 200 VK-----DNKQRTKKKPPKPPVNBAGSLDNGDFKVTTPDSTQHNK 245
 386 VQWALDIEIAARKLLEGEBCIGFGSP-----PSLLEGLPKI--PSIST--HIK 432
 246 VSTSPKITTAKPINDPSPILPNSDTSKETSITVKNKETTVEKTTTNKQSTDGKETT 305
 433 VKSEEMIKVVE-----KSEKETVIVEGQTEIRVTEGVTEEDKBAQOGEH 480
 306 SAKETOSTIKTSAKOLAPTSKVLAKPTKRAETTGPAITTKPEPTTPKREPASTTPKE 365
 481 AEEGEKEEELIAATSPAEBAASPEKETSIVVEEAKSPGEAKSPGEAKSPA---EA 536
 366 PPTTIKAPPTPKPEAPPTTKASAPTTKPEP--APTTKEPAPTTPKPEAPTTKPEP--- 420
 537 KSPGSAKS--PGSAKSPGEAKSPAEPKSPAEPKSPAEPKSPA--EKSRA--TVKSPGEAK 591
 421 APTTKASAPTTKPEPA---PTTPKPA---PTTPKPEA---BTTPKEP---TPTTP 464
 592 SPSEAKS--PAEAKSPAEEKSPAEEKSPAEEKSPAEEKSPAEEKSPAEEKSPA 650
 465 KEPAPTTKPEAPTTKPEPA---PTAAPP---APTTKPEAPTTKPEAPTTKPEP 516
 651 KSPA--EKSAPAEKSPAEPKSPAEPKSPAEPKSPAEPKSPA--KSPA---EVSP 701
 517 TTPKPEA---PTTKASAPTTKPEAPTTKSAPTTKPE--SPPTTKPEA---PTTPKE 567
 702 AEAKSPAEPKSPGEAKS--PAVKSAPAEKSPAAPVSPGEAKSPGSAKSPAEEKSPA 760
 568 PAPTPKKAPPTPKPEAPTTKPEAPTTKPKKAPPAKPEAPTTKPEAPTTKPKK 627
 761 PIEVKSPEKATPVGEAKSPAEEKSPAEPKSPVKEKIKPPAEKSPKA--KSPVGEAK 819
 628 TPKEKAPPTPKPEAPTTPEELAPTTPEPTTPPEE--PAPTPKAAAPNTPKPEAPT-- 684
 820 PEKAPKPDVKSPEAOTVQEBATVPDIR--PPOVKSPEAKKSPKSP--KEEAKTSK 875
 685 --PKPEAPPTP-----KEPAPTTKETAPTTPGTAPTTKPEAPTTKPKKAPKELAP 735
 876 VAPKEVKSPEVKEVAKPEPKKYBEKTLPTPKESKSDAPKAPKPEKKE 935
 936 TPEKPKDSTAEEKKEAGEKKKAVASBEFPKAGVKEEA--KKEKETETKTAEDTK 993
 736 TTKKEPTSTSD--KPAPTPKGAAPTTKPEAPTTKPEAPTTKPGTAPTTKPEAPTT 793
 994 AEKPS--KPTETKP-----KKEKPPAPK----- 1017
 854 TSEVSTPTTKPEPTTIHKSDESTPELSAEPKALENSPEKPGVPTTKPATKPEMTT 913
 1018 -----KOTKEKTESKPEK-----PK-MAKVKEDDSLSKEP--SKPKTEK 1059
 914 TAKDKTERDLRTTPETTAAPKMTKETATTTK 947
 1060 AEKSSSTQKESQPE-----KTTEDKATGK 1087

Mon Dec 8 09:50:45 2003

us-09-556-246-1.rsp

Page 22

Search completed: December 8, 2003, 09:35:50
Job time : 28 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd

OM protein - protein search, using sw model

Run on: December 8, 2003, 09:19:41 ; Search time 54 Seconds
(without alignments)
4126.891 Million cell updates/sec

Title: US-09-556-246-1
 Page: 7526

Perfect score: 7526
Sequence: 1 MAMKTLPIYLLLLSVFIQ.....ARAITRSGQTLSKWMYNCP 1404

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: . 1107863 beqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

```
Post-processing:  Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries
```

Database :

A_Geneseq_15Jun03:*

1: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1980.DAT *

2: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1981.DAT *

3: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1982.DAT *

4: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1983.DAT *

5: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1984.DAT *

6: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1985.DAT *

7: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1986.DAT *

8: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1987.DAT *

9: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1988.DAT *

10: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1989.DAT *

11: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1990.DAT *

12: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1991.DAT *

13: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1992.DAT *

14: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1993.DAT *

15: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1994.DAT *

16: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1995.DAT *

17: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1996.DAT *

18: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1997.DAT *

19: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1998.DAT *

20: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1999.DAT *

21: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT *

22: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT *

23: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2002.DAT *

24: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2003.DAT *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	7526	100.0	1404	13	AA826649	NSF precursor. Sy
2	7526	100.0	1404	22	AA860568	Human megakaryocy
3	7526	100.0	1404	22	AA829773	Human megakaryocy
4	7523	100.0	1415	22	AAJ32462	Novel human secret
5	6950	92.3	1289	22	AA824322	Human EST encoded
6	3484	46.3	902	22	AA829778	Human NSF-derived
7	2929	38.9	551	22	ABU53552	Human testis-deriv
8	2920	38.8	546	22	ABU53552	Human testis-deriv
9	2850.5	37.9	538	23	AA018834	Human cartilage supe

10	2757	36.6	51.2	22	ABUS3254	Human testes-deriv
11	2197	29.2	45.2	16	AA880041	Human megakaryocyt
12	1707.5	22.7	42.2	22	AA560559	Bovine MSR ortholo
13	1545	20.5	29.2	22	AAU11261	Human HAP0 polypep
14	1188	15.8	51.79	22	AA2424516	C89sp predicted am
15	1188	15.8	51.79	24	ABP55355	Human colon tumour
16	1097	14.6	21.4	22	ABUS3255	Human testes-deriv
17	1012.5	13.5	71.7	22	ABUS3144	Human testes-deriv
18	981	13.0	76.3	21	AA338992	Arabidopsis thalia
19	968.5	12.9	188	23	AA018833	5' cartilage super
20	950	12.6	1664	29	AAW43106	C. thermocellum O1
21	946.5	12.6	1049	12	AB861364	Drosophila melanog
22	927.5	12.3	778	22	ABUS3143	Human testes-deriv
23	927	12.3	770	22	ABUS3141	Human testes-deriv
24	902	12.0	1795	22	AB869806	Drosophila melanog
25	900.5	12.0	71.7	22	ABUS3145	Human testes-deriv
26	900.5	12.0	71.7	22	ABUS3146	Human testes-deriv
27	900.5	12.0	71.7	22	ABUS3147	Human testes-deriv
28	900.5	12.0	71.7	22	ABUS3148	Human testes-deriv
29	900.5	12.0	71.7	22	ABUS3149	Human testes-deriv
30	900.5	12.0	71.7	22	ABUS3150	Human testes-deriv
31	900.5	12.0	71.7	22	ABUS3151	Human testes-deriv
32	890	11.8	74.5	22	ABUS3142	Human testes-deriv
33	889	11.8	69.2	22	ABUS3152	Human testes-deriv
34	848.5	11.3	69.5	22	ABUS3154	Human testes-deriv
35	818.5	10.9	64.5	22	ABUS3152	Human testes-deriv
36	805	10.7	13.25	22	AB849735	Human liver peptid
37	805	10.7	13.25	22	AB839735	Peptide #2376 enco
38	805	10.7	13.25	22	AB834897	Peptide #2403 enco
39	805	10.7	13.25	22	AB8200314	Protein #2313 enco
40	805	10.7	13.25	22	AA855707	Human brain expres
41	805	10.7	13.25	22	AAAM68085	Human bone marrow
42	805	10.7	13.25	22	AAAM36645	Human #2327 enco
43	805	10.7	13.25	23	ABG37612	Human peptide enco
44	782.5	10.4	21.12	23	AB860403	Drosophila melanog
45	776	10.3	39.5	22	ABUS3160	Human testes-deriv

RESULT	1
ID	AAR26049 standard; Protein; 1404 aa.
XX	
AC	AAR26049;
XX	
DT	25-MAR-2003 (updated)
DT	02-FEB-1993 (first entry)
XX	
DE	MSF precursor.
XX	
KM	Megakaryocyte colony stimulating factor; secretion signal; meg-CSF
KX	stability; proteolytic cleavage; adhesion; alternative splicing.
XX	
OS	Synthetic.
FH	Key
FT	Region
FT	Location/Qualifiers
FT	1..26
FT	/label= Exon_I
FT	26..67
FT	/label= Exon_II
FT	67..107
FT	/label= Exon_III
FT	107..157
FT	/label= Exon_IV
FT	157..200
FT	/label= Exon_V
FT	200..1141
FT	/label= Exon_VI
FT	141..1166
FT	/label= Exon_VII
FT	1166..1212
FT	Region

THIS PAGE BLANK (USPTO)

FT /label= Exon_VIII
FT Region 1213..1266
FT /label= Exon_IX 1266..1331
FT Region /label= Exon_X 1331..1373
FT Region /label= Exon_XI 1373..1404
FT Region /label= Exon_XII

PN M09213075-A1.
XX 06-AUG-1992.
PD
XX 17-JAN-1992; 92MO-US00433.
PP
XX 18-JAN-1991; 91US-0643502.
PR 10-SEP-1991; 91US-0757022.
R
X (GENY) GENETICS INST INC.
A
X Clark SC, Gesner TG, Hewick RM, Jacobs K, Turner K;
X WPI, 1992-284660/34.
R N-PSDB; AA027223.
XX
XX New human mega-karyocyte stimulating factors - for treating
PT immune deficiencies, cancer, exposure to radiation or drugs,
PT bacterial and viral infections, etc.
XX
XX Claim 1, 2 and 3; Fig 1; 87pp; English.

XX The sequence given is a full length translation from the megakaryocyte
CC stimulating factor (MSF) precursor. The sequence covered by exons II,
CC III and IV encodes megakaryocyte stimulating factor (MSF). This
CC sequence is modified by the addition of an N-terminal sequence encoding
CC a secretory leader, an initiating methionine preceding exon II and a
CC terminating codon following exon IV. The cDNA sequence given contains
CC sequences derived from human megakaryocyte colony stimulating factor
CC (meg-CSF). Exon I contains the initiating methionine, and encodes a
CC classical mammalian protein secretion signal sequence. The sequence
CC encoding the original meg-CSF includes exons II-IV and is thought to
CC terminate in the region between amino acid residues 134 - 147. The
CC primary transcript of this gene may be cleaved in different ways to
CC yield a family of mRNA's each encoding a different MSF protein. Exons
CC V and VI are thought to be related to the activity of the factor and
CC are also implicated in the stability, folding and processing of the
CC molecule. These exons are also thought to play a role in the observed
CC synergy of MSF with other cytokines. Exons V - XII are believed to be
CC implicated in the processing or folding of the appropriate structure of
CC the resulting factor, ie. one or more of these exons may contain
CC sequences which direct proteolytic cleavage, adhesion, organization of
CC the cellular matrix or extracellular matrix processing. Both naturally
CC occurring and non-naturally occurring MSF's may be characterized by
CC various combinations of alternatively spliced exons from this sequence,
CC with the exons spliced together in differing orders to form different
CC members of the MSF family.
CC (Updated on 25-MAR-2003 to correct PN field.)
XX
XX

SO Sequence 1404 AA;
Query Match 100.0%; Score 7526; DB 13; Length 1404;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1404; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAMKTLPIYLILLSVPIIOVSSQDSSCAGROEGYSRDATNCQVNCCHNECCPDF 60
DB 1 MAMKTLPIYLILLSVPIIOVSSQDSSCAGROEGYSRDATNCQVNCCHNECCPDF 60
QY 61 KRVCTABISCKRCPESEFERGECDCDAQCKKYDKCCPDYESFCAEVANPTSPSSKAP 120
DB 61 KRVCTABISCKRCPESEFERGECDCDAQCKKYDKCCPDYESFCAEVANPTSPSSKAP 120

QY 121 PPSGASQITIKSTTKASPRPKKKTKKVIIESREITEHSVSENOSSSSSSSSSTIM 180
DB 121 PPSGASQITIKSTTKASPRPKKKTKKVIIESREITEHSVSENOSSSSSSSSSTIM 180
QY 181 KIKSSKSNAAANELOKQKVONKNGRTKKKPTPKPVVDEAGSLDNGDFVUTPDIST 240
DB 181 KIKSSKSNAAANELOKQKVONKNGRTKKKPTPKPVVDEAGSLDNGDFVUTPDIST 240
QY 241 TOHNVSTSPKLTAKPINRPSLPNSDTSKETSILVNNKETTVETKETITNNKQSTDG 300
DB 241 TOHNVSTSPKLTAKPINRPSLPNSDTSKETSILVNNKETTVETKETITNNKQSTDG 300
QY 301 KEKTTSAKETOSEIKTSADLAFTSKVLAAPPKAETTTKGPALTTPKSEPTTTKEPS 360
DB 301 KEKTTSAKETOSEIKTSADLAFTSKVLAAPPKAETTTKGPALTTPKSEPTTTKEPS 360
QY 361 TTPKEPTPTTISAPTTPEKAPATTTKSAPTTPEKAPATTPEKAPATTPEKAP 420
DB 361 TTPKEPTPTTISAPTTPEKAPATTTKSAPTTPEKAPATTPEKAPATTPEKAP 420
QY 421 APTTTSAPTTPEKAPATTPKKAPATTPEKAPATTPEKAPATTPEKAPATTPEK 480
DB 421 APTTTSAPTTPEKAPATTPKKAPATTPEKAPATTPEKAPATTPEKAPATTPEK 480
QY 481 EPAPTPAPKKAPATTPEKAPATTPEKAPATTPEKAPATTPEKAPATTPEKAP 540
DB 481 EPAPTPAPKKAPATTPEKAPATTPEKAPATTPEKAPATTPEKAPATTPEKAP 540
QY 541 TTKSAPTTPEKSPPTTKEKAPATTPEKAPATTPEKAPATTPEKAPATTPEKAP 600
DB 541 TTKSAPTTPEKSPPTTKEKAPATTPEKAPATTPEKAPATTPEKAPATTPEKAP 600
QY 541 TTKSAPTTPEKSPPTTKEKAPATTPEKAPATTPEKAPATTPEKAPATTPEKAP 600
DB 541 TTKSAPTTPEKSPPTTKEKAPATTPEKAPATTPEKAPATTPEKAPATTPEKAP 600
QY 601 APTAPKAPATTPKSTAPTTPKCLPTTPEKLAAPTTPEKAPATTPEKAPATTPE 660
DB 601 APTAPKAPATTPKSTAPTTPKCLPTTPEKLAAPTTPEKAPATTPEKAPATTPE 660
QY 661 PEPAPATTKAAANTPKAPATTPEKAPATTPEKAPATTPEKAPATTPEKAPATT 720
DB 661 PEPAPATTKAAANTPKAPATTPEKAPATTPEKAPATTPEKAPATTPEKAPATT 720
QY 721 APTTPKAPKELAPTTKEPTSTSDKAPATTPGTAPTTPEKAPATTPEKAPATT 780
DB 721 APTTPKAPKELAPTTKEPTSTSDKAPATTPGTAPTTPEKAPATTPEKAPATT 780
QY 781 TAPTTKEBAPTTPKKAPKELAPTTKGPTSTSDKAPATTPKSTAPTTPEKAPATT 840
DB 781 TAPTTKEBAPTTPKKAPKELAPTTKGPTSTSDKAPATTPKSTAPTTPEKAPATT 840
QY 841 KPAPTTPEPTPTTSVSTPTTKEPTTHKSPDESTPELSAEPKALENSPEKGVPT 900
DB 841 KPAPTTPEPTPTTSVSTPTTKEPTTHKSPDESTPELSAEPKALENSPEKGVPT 900
QY 901 TTPPATKEMTTAKDKTERDLRTTPETTAAAPMTGETATTTKETTESKITATTTOV 960
DB 901 TTPPATKEMTTAKDKTERDLRTTPETTAAAPMTGETATTTKETTESKITATTTOV 960
QY 961 TSTTTODTTPFKITLKTITLAPKYTTTKITITTEINMKPEETAAPKORATNSKATTPK 1020
DB 961 TSTTTODTTPFKITLKTITLAPKYTTTKITITTEINMKPEETAAPKORATNSKATTPK 1020
QY 1021 POKPTAPPKKPTSTKPKTMKPRVRKPTPTPRKXTSTNPELNPSTRIAEALQTTTRN 1080
DB 1021 POKPTAPPKKPTSTKPKTMKPRVRKPTPTPRKXTSTNPELNPSTRIAEALQTTTRN 1080
QY 1081 QTPNSKLVEVNPSESDAGAGETPHMLLRPHVPEVTPPMDVYLPRVNOGIINPMLS 1140
DB 1081 QTPNSKLVEVNPSESDAGAGETPHMLLRPHVPEVTPPMDVYLPRVNOGIINPMLS 1140
QY 1141 DETNINCGKPVQGLTTLRNGTLVAFRGHYFNNLSPPSPPARITVEVGIDSPIDTFT 1200
DB 1141 DETNINCGKPVQGLTTLRNGTLVAFRGHYFNNLSPPSPPARITVEVGIDSPIDTFT 1200
QY 1201 RNCBCKTFFPFDSQVYRFTNLIKQAGYKPIPFKRGGLTGOI VALASTAKKNPESYV 1260

THIS PAGE BLANK (USPTO)

DB 1201 RCGCEGTFFPKDSQVWRFTNDIKDAGYKPIFGFGGLTGQVAAALSTAKYKMWESVY 1260
QY 1261 PFRGGSIOQYIYKQBPVQCPGRRPALNTPYVGMQVRRRFRPAIGSQHTTRIQY 1320
D 1261 PFRGGSIOQYIYKQBPVQCPGRRPALNTPYVGMQVRRRFRPAIGSQHTTRIQY 1320
QY 1321 SPARLAVQDGVLMNEKVSILMRGLPNVYTSALSLPNIRKPDGYDYAFSKQYVNDY 1380
DB 1321 SPARLAVQDGVLMNEKVSILMRGLPNVYTSALSLPNIRKPDGYDYAFSKQYVNDY 1380
QY 1381 PSRTARALITRSQGLTSKWNVNC 1404
DB 1381 PSRTARALITRSQGLTSKWNVNC 1404
RESULT 2
AAB60568
ID AAB60568 standard; Protein, 1404 AA.
AC AAB60568;
XX
DT 27-APR-2001 (first entry)
XX
DB Human megakaryocyte stimulating factor (MSF, CACP).
XX
KW Human, CACP protein, campodactylly-arthropathy-coxa vara-pericarditis;
KW MSF; megakaryocyte stimulating factor; synovial lubricant;
KW chromosome 1q25-31; osteoarthritis; joint lubrication; osteopathic;
KW anticharctic.
XX
OS Homo sapiens.
XX
PN M0200107068-A1.
XX
PD 01-FEB-2001.
XX
PE 21-JUL-2000; 2000MO-US20002.
XX
PR 23-JUL-1999; 9905-0145328.
PR 19-JUL-2000; 200005-0145328.
XX
PA (UYCA-) UNIV CASE WESTERN RESERVE.
XX
PI Matman M.
XX
DR WPI; 2001-182721/18.
XX
PT New composition comprising the campodactylly-arthropathy-coxa
PT vara-pericarditis protein in combination with an anesthetic, useful for
PT treating osteoarthritis, or as lubricants of tissue and joints -
XX
PS Example 1; Page -: 34pp; English.
XX
CC The invention relates to a method of treating osteoarthritis via the
CC administration of a composition comprising the campodactylly-arthropathy-
CC coxa vara-pericarditis (CACP) protein, or portions of the CACP protein.
CC The composition may further comprise a local anesthetic. The composition
CC of the invention may be administered via intra-articular or intravenous
CC injection. The human CACP protein is identified in the invention as
CC being megakaryocyte stimulating factor (MSF). The gene encoding
CC CACP protein (MSF) is located on chromosome 1q25-31, and mutations in
CC this gene are responsible for the heritable disorder campodactylly-
CC arthropathy-coxa vara-pericarditis, in which patients have synovial
CC hyperplasia without evidence of inflammation. CACP protein (MSF)
CC acts as a synovial lubricant, and can be used to lubricate tissue and
CC joints in the treatment of osteoarthritis. The composition may be
CC applied to reduce the symptoms of osteoarthritis (e.g., joint pain,
CC loss of range of movement or joint damage). The present sequence
CC represents human megakaryocyte stimulating factor (MSF, CACP protein).
CC Note: This sequence is not given in its entirety in Figure 4 of the
CC specification, although a Genbank accession number was given. This
CC sequence was therefore obtained from Genbank (U70316).

XX	Sequence	1404 AA;	SQ
QY	Query Match	100.0%; Score 7526; DB 22; Length 1404;	
DB	Best Local Similarity	100.0%; Pred. No. 0;	
DB	Matches 1404; Conservative	0; Mismatches 0; Indels 0; Gaps 0;	
QY	1	MAAKTLPIYLILLSVFVLIQVVSQDLSSCAGRCCEGYSRDATCNDYVCOHYMECCPDF 60	
DB	1	MAAKTLPIYLILLSVFVLIQVVSQDLSSCAGRCCEGYSRDATCNDYVCOHYMECCPDF 60	
QY	61	KAVCTAELSCKRCESFERGECDDAQCCKYDKCCPYESFCARVHNPSPSSKAP 120	
DB	61	KAVCTAELSCKRCESFERGECDDAQCCKYDKCCPYESFCARVHNPSPSSKAP 120	
QY	121	PPSGASQITKSTTKSPKPNKKTKVLESSEITHEHVSNOSSSSSSSSSTIM 180	
DB	121	PPSGASQITKSTTKSPKPNKKTKVLESSEITHEHVSNOSSSSSSSSSTIM 180	
QY	181	KIKSSKNSANRELQKLVKONKKNRTKKKPPVVDGASGLNDGPKVTTDTST 240	
DB	181	KIKSSKNSANRELQKLVKONKKNRTKKKPPVVDGASGLNDGPKVTTDTST 240	
QY	241	TOHNKVSTSPKITTAKPINRPSLPNSDTSKETSITNKKETTVETKETTNNKOTSDG 300	
DB	241	TOHNKVSTSPKITTAKPINRPSLPNSDTSKETSITNKKETTVETKETTNNKOTSDG 300	
QY	301	KERTTSKAKTOSIEKTSADOLAPTSKVLAKPTPKAETTTKGPALTTPKBPTTPKBPAS 360	
DB	301	KERTTSKAKTOSIEKTSADOLAPTSKVLAKPTPKAETTTKGPALTTPKBPTTPKBPAS 360	
QY	361	TPKBPPTTTIKSAPTTTPKEBAPATTTKSAPTTTPKEBAPATTTPKEBAPATTTPKEB 420	
DB	361	TPKBPPTTTIKSAPTTTPKEBAPATTTKSAPTTTPKEBAPATTTPKEBAPATTTPKEB 420	
QY	421	APTTSKAPTTTPKEBAPTTPKKAPATTTPKEBAPTTTPKEBAPTTTPKEBAPTTTPKEB 480	
DB	421	APTTSKAPTTTPKEBAPTTPKKAPATTTPKEBAPTTTPKEBAPTTTPKEBAPTTTPKEB 480	
QY	481	EPAPTPPKKAPATTTPKEBAPTTTPKEBAPTTTPKEBAPTTTPKEBAPTTTPKEBAPTT 540	
DB	481	EPAPTPPKKAPATTTPKEBAPTTTPKEBAPTTTPKEBAPTTTPKEBAPTTTPKEBAPTT 540	
QY	541	TTKSAPTTTPKEBAPTTTPKEBAPTTTPKEBAPTTTPKEBAPTTTPKEBAPTTTPKEB 600	
DB	541	TTKSAPTTTPKEBAPTTTPKEBAPTTTPKEBAPTTTPKEBAPTTTPKEBAPTTTPKEB 600	
QY	601	APTAPKEBAPTTTPKEBAPTTTPKEBAPTTTPKEBAPTTTPKEBAPTTTPKEBAPTTTP 660	
DB	601	APTAPKEBAPTTTPKEBAPTTTPKEBAPTTTPKEBAPTTTPKEBAPTTTPKEBAPTTTP 660	
QY	661	PEBPAPTTTPKAAAPNPKBAPTTTPKEBAPTTTPKEBAPTTTPKEBAPTTTPKEBAPTT 720	
DB	661	PEBPAPTTTPKAAAPNPKBAPTTTPKEBAPTTTPKEBAPTTTPKEBAPTTTPKEBAPTT 720	
QY	721	APTTPKAPKAPKELAPTTTPKSTSDKAPTTPKGAPTTTPKEBAPTTTPKEBAPTTTPKG 780	
DB	721	APTTPKAPKAPKELAPTTTPKSTSDKAPTTPKGAPTTTPKEBAPTTTPKEBAPTTTPKG 780	
QY	781	TAPTTPKEBAPTTTPKAPKELAPTTTPKSTSDKAPTTPKGAPTTTPKEBAPTTTPKEBAP 840	
DB	781	TAPTTPKEBAPTTTPKAPKELAPTTTPKSTSDKAPTTPKGAPTTTPKEBAPTTTPKEBAP 840	
QY	841	KPAPTTPETPTPTST 900	
DB	841	KPAPTTPETPTPTST 900	
QY	901	TKTPATKPEMTTAKOKTERDLRTTETTTAPKATKATATTEKTSKITTATTVQ 960	
DB	901	TKTPATKPEMTTAKOKTERDLRTTETTTAPKATKATATTEKTSKITTATTVQ 960	
QY	961	TSITTQTTTTPKITTAKTTTAAKVTYTTKTTTTEIMNKPESTAKPKDRAVNSKATTPK 1020	
DB	961	TSITTQTTTTPKITTAKTTTAAKVTYTTKTTTTEIMNKPESTAKPKDRAVNSKATTPK 1020	

THIS PAGE BLANK (USPTO)